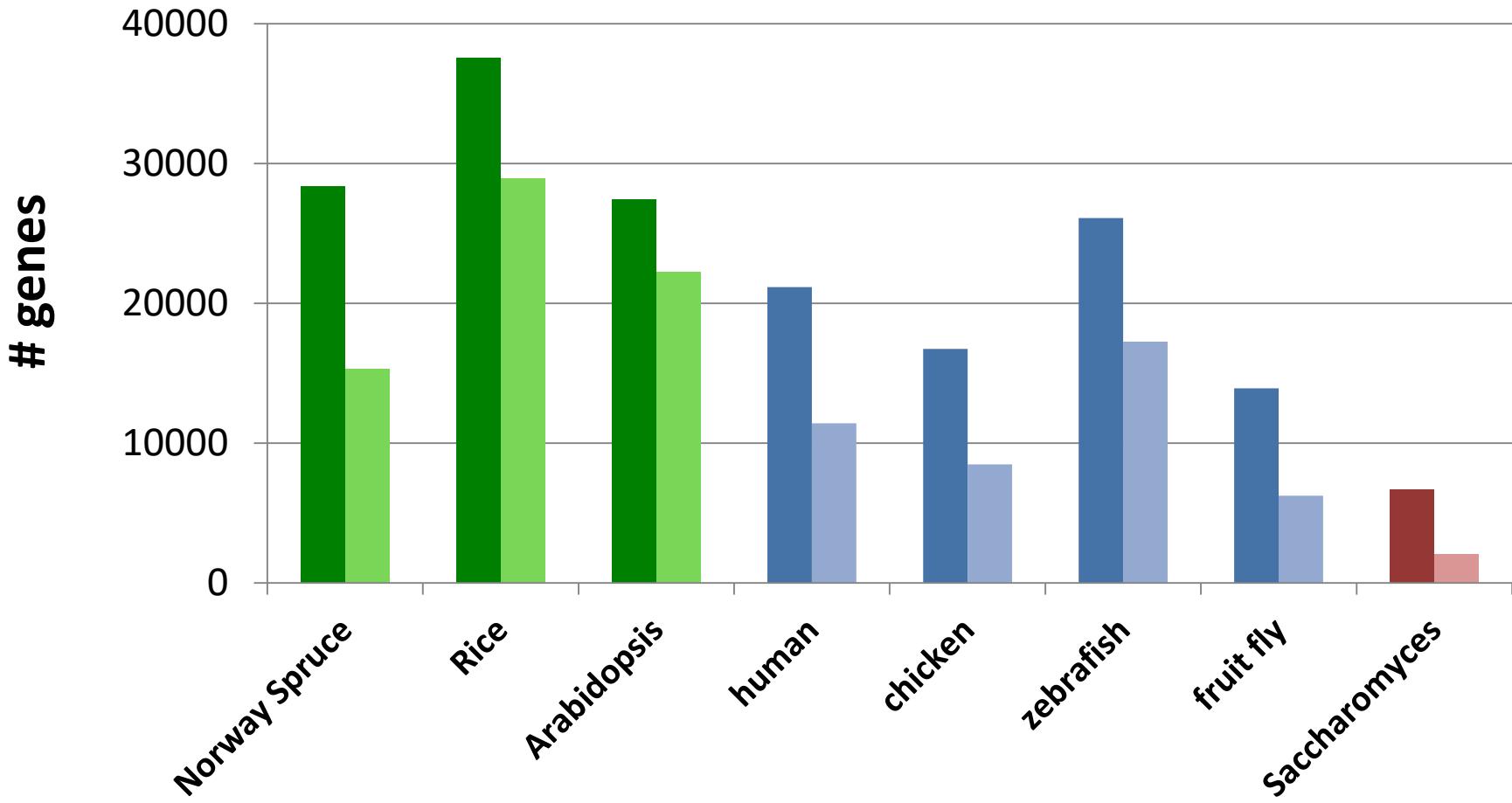


# **Analysis of Pinaceae gene family evolution suggests accelerated gene turnover in pine trees**

**Claudio Casola**

Department of Ecosystem Science and Management

# Duplicated genes are common in genomes



Dark color = all genes

Light color = duplicated genes

# Comparative genomics in forest trees

“The understanding of gene function in forest tree species is challenged by the lack of standard reverse-genetic tools routinely used in other systems [...] and by the long generation times. Thus, comparative genomics becomes the more powerful approach to understanding gene function in trees.”

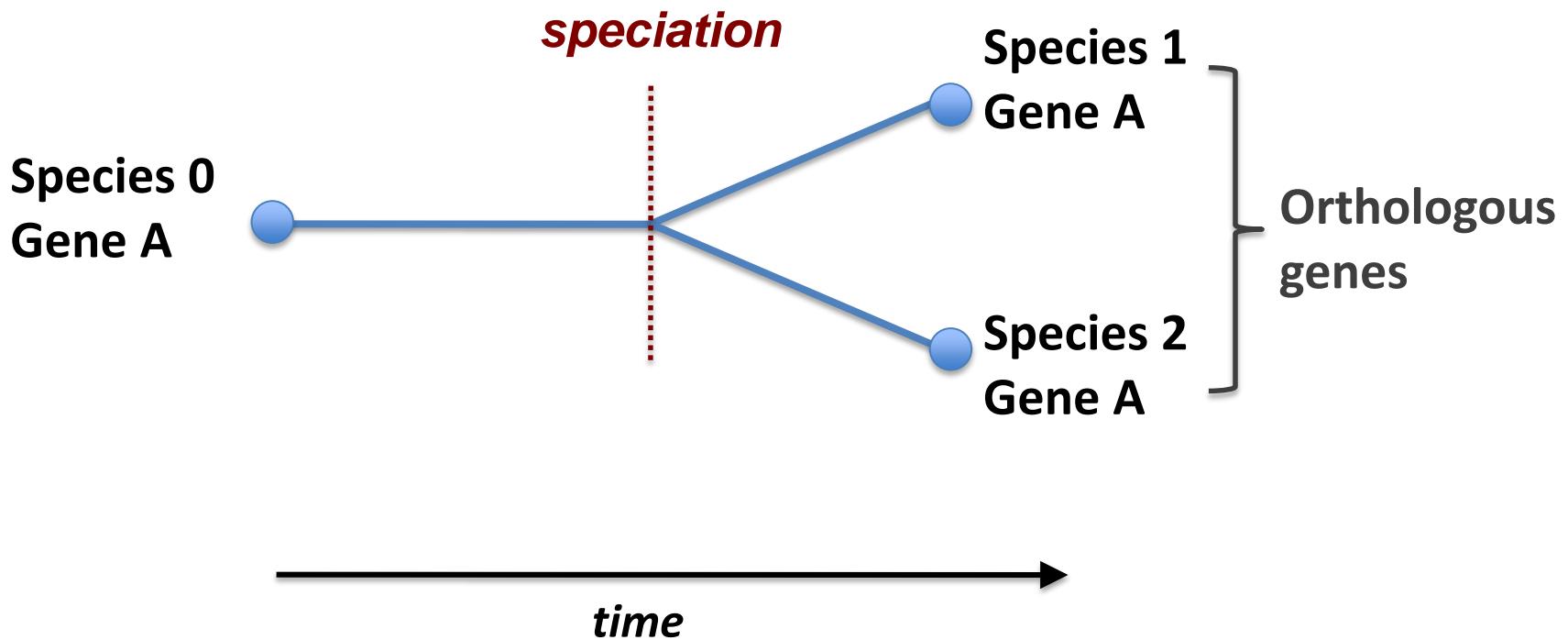
*Neale et al. “Open access to tree genomes: the path to a better Forest” Genome Biology 2013*

# Orthologous and paralogous genes

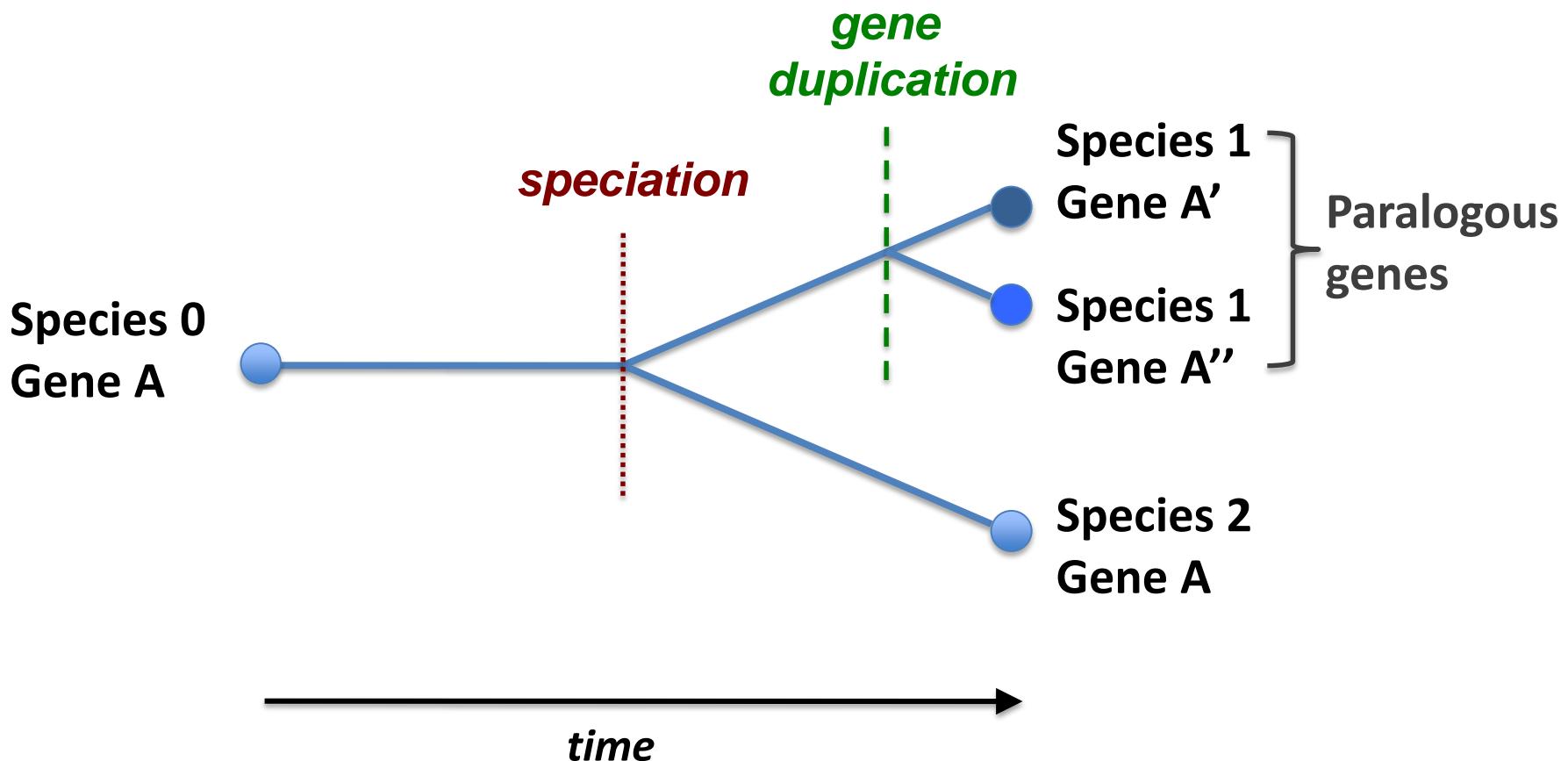
Species 0  
Gene A



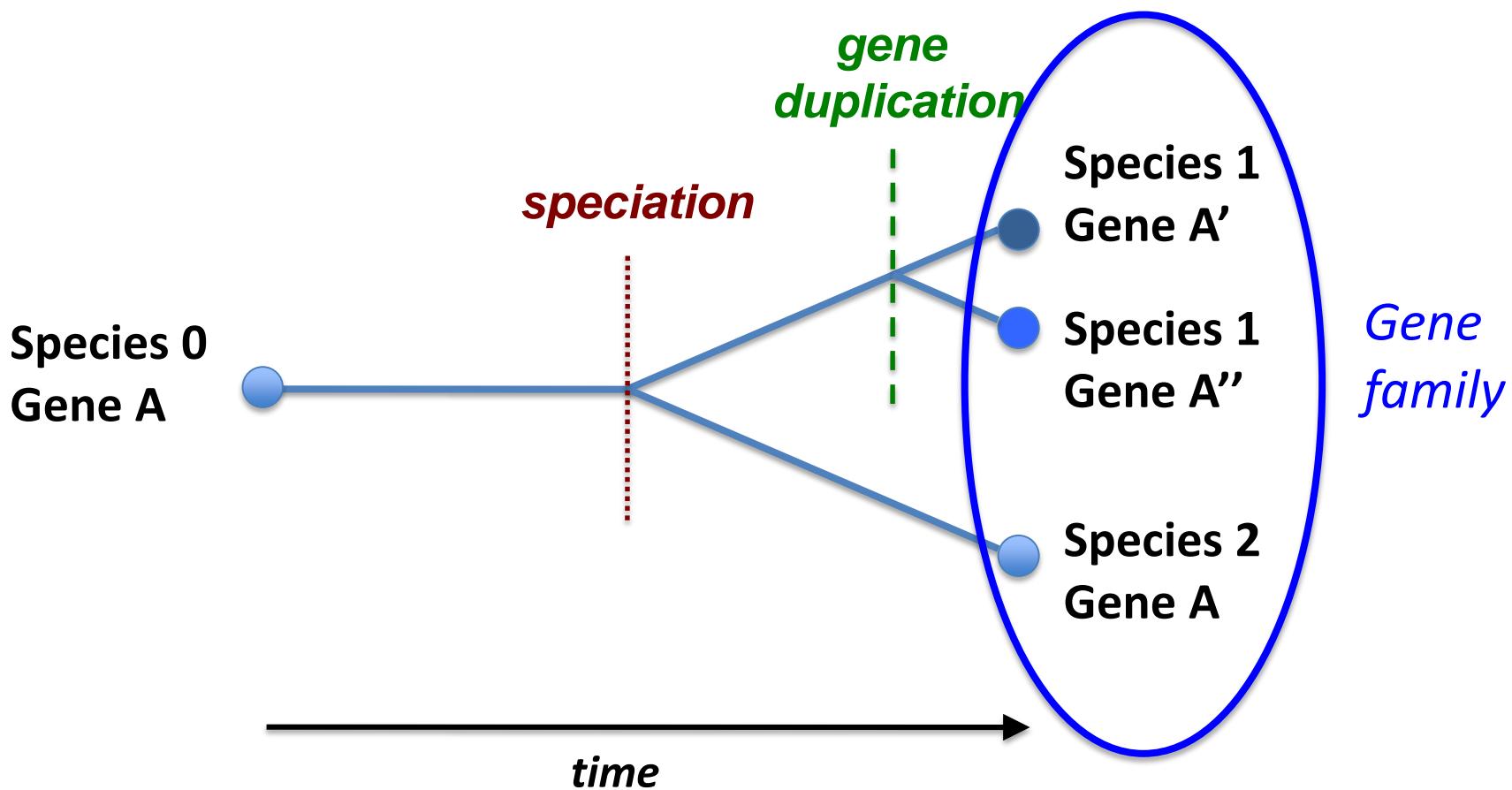
# Orthologous and paralogous genes



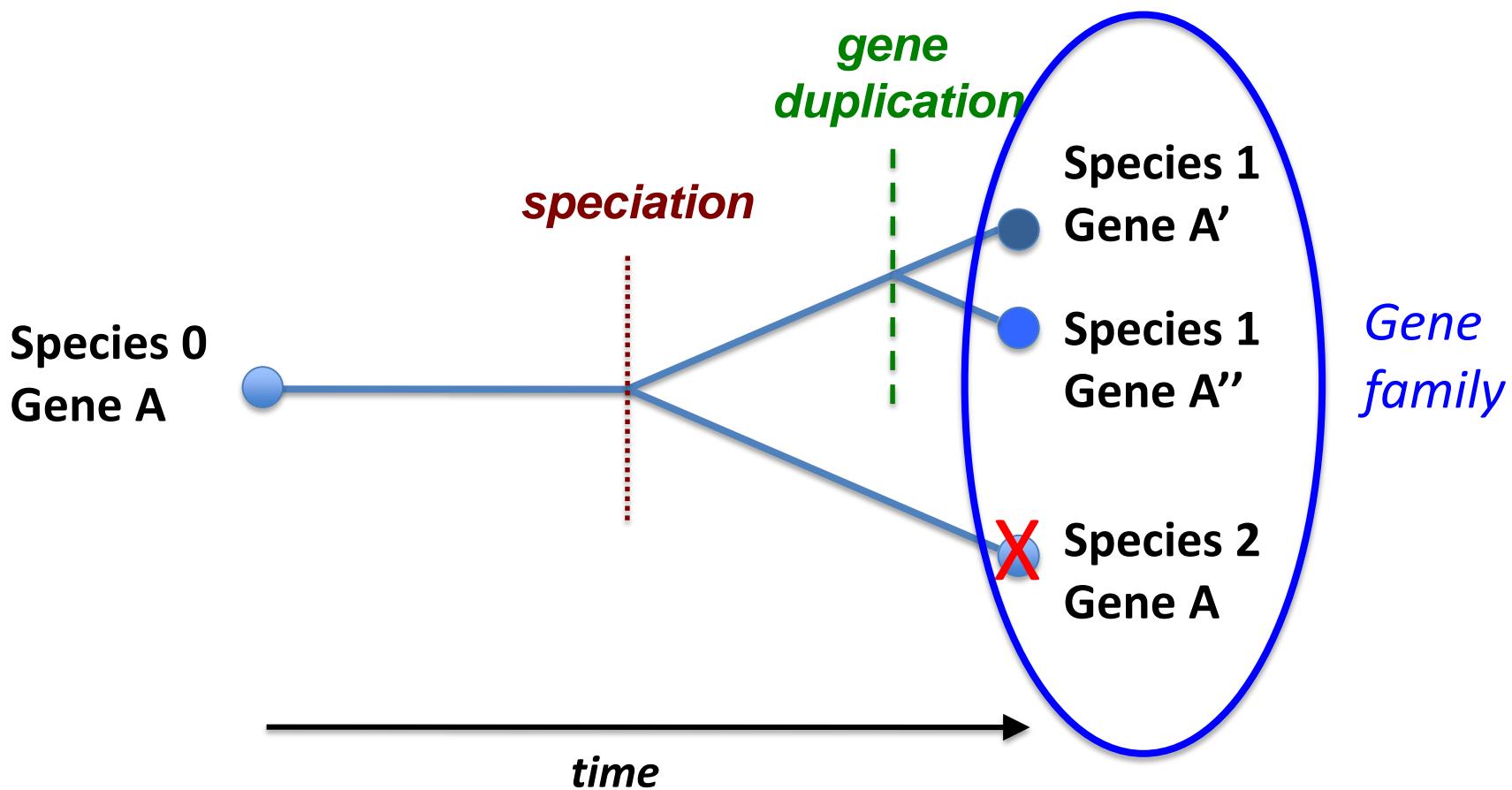
# Orthologous and paralogous genes



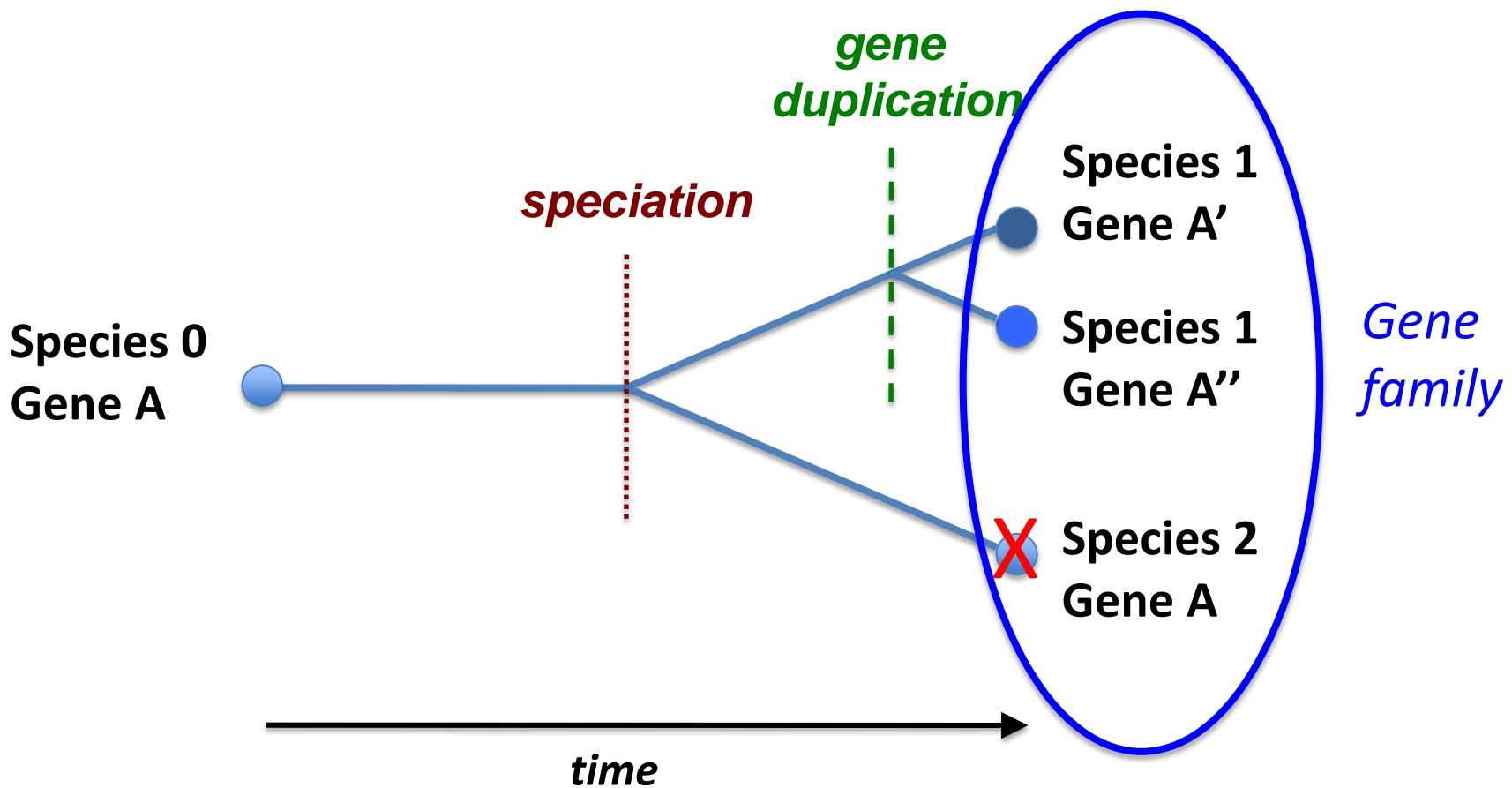
# Orthologous and paralogous genes



# Orthologous and paralogous genes



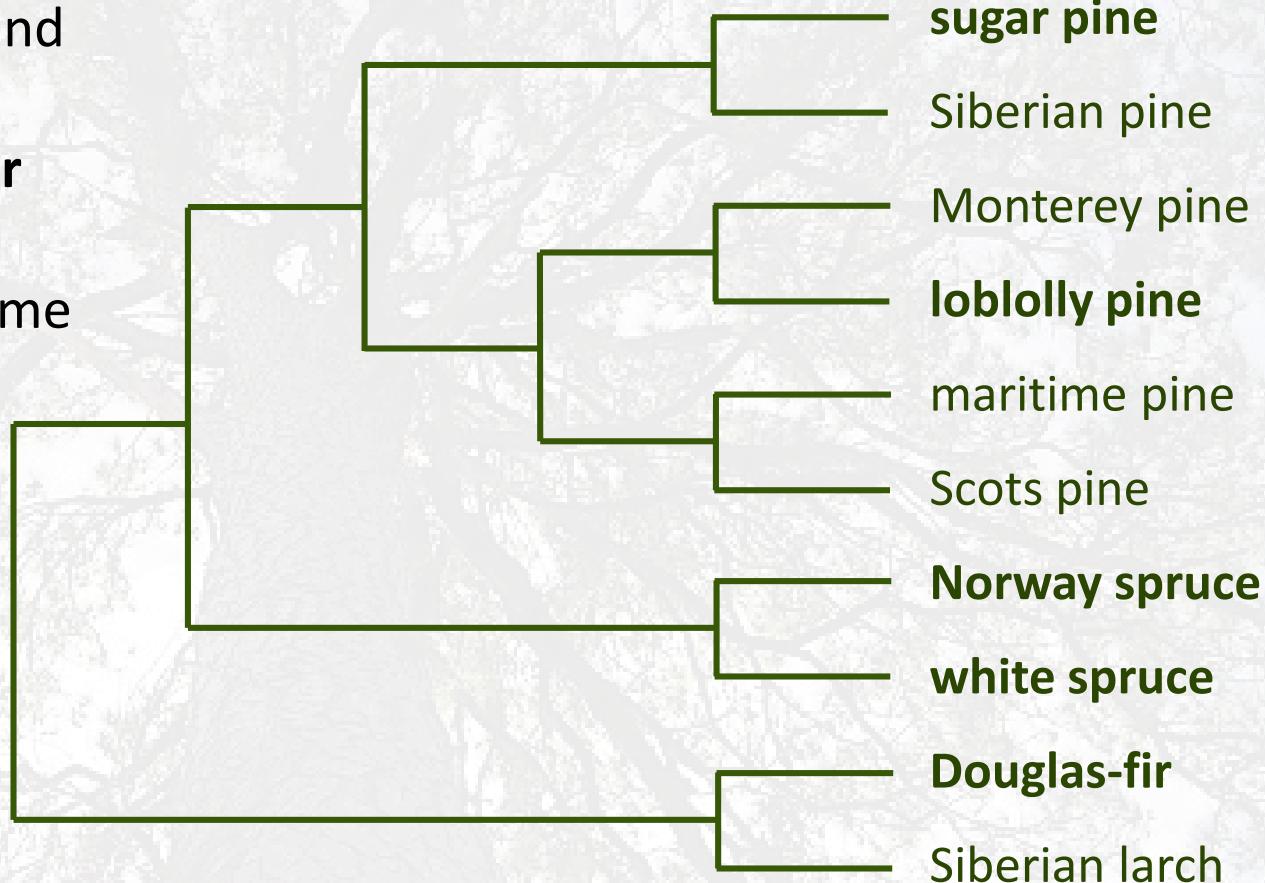
# Orthologous and paralogous genes



**Gene turnover:** the combination of gene duplications and gene losses in a given lineage through time == Change in gene family size

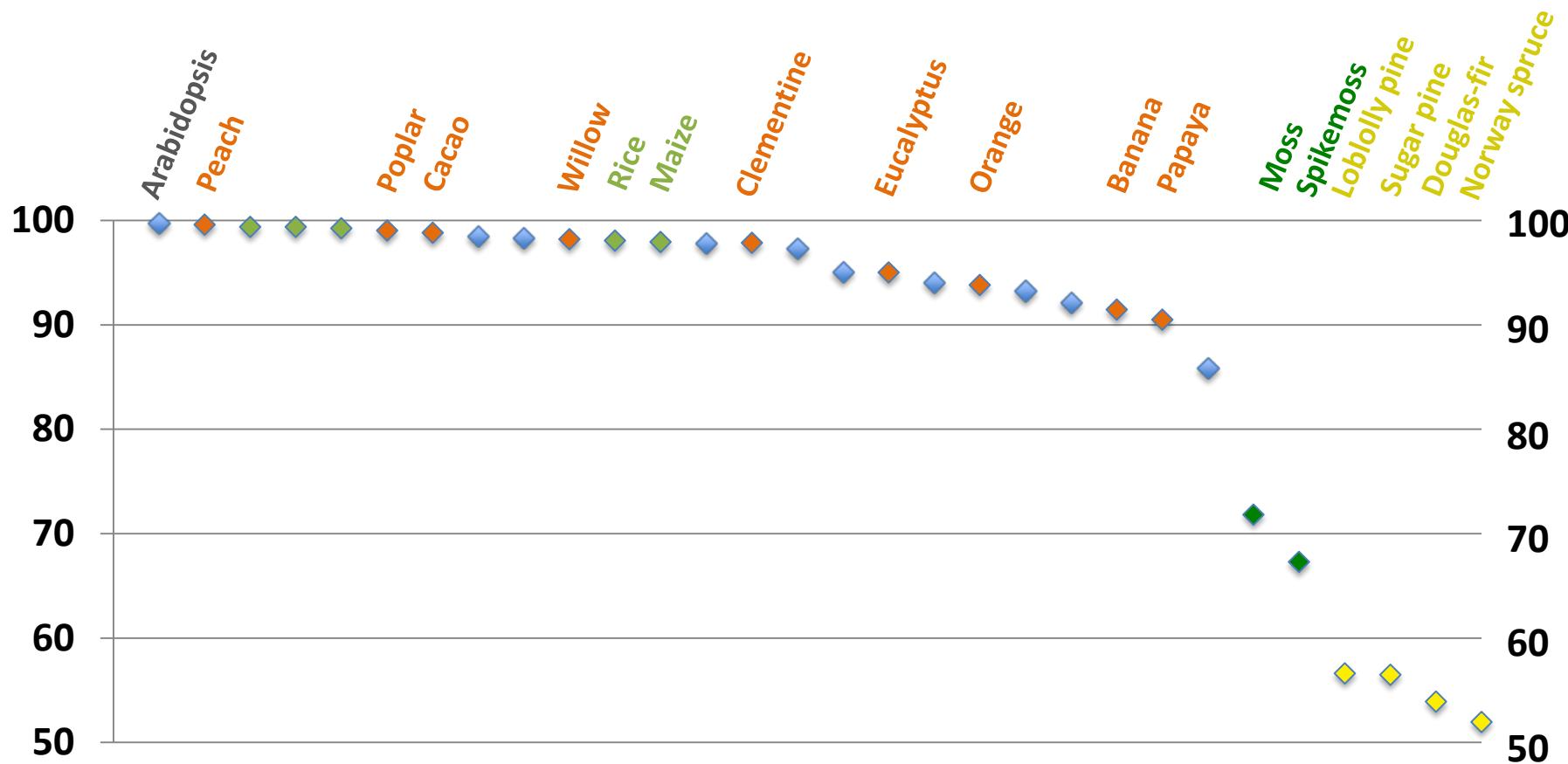
# Gene evolution in Pinaceae

In progress and  
**complete**  
**(published or  
submitted)**  
conifer genome  
sequencing  
projects



Phylogeny from <http://tolweb.org/Pinaceae/>

# Completeness of gene annotation in seed plants



Analysis with *BUSCO*. Proportion of complete and fragmented genes among 1440 1-to-1 embryophyta orthologous genes

# Outline

## ◎ Different rates of gene turnover in Pinaceae

1. Issues with gene fragments (false gene duplications)
2. Issues with gene duplications vs. gene losses
3. Issues with species phylogeny

## ◎ Rates of gene turnover in Pinaceae vs. angiosperms

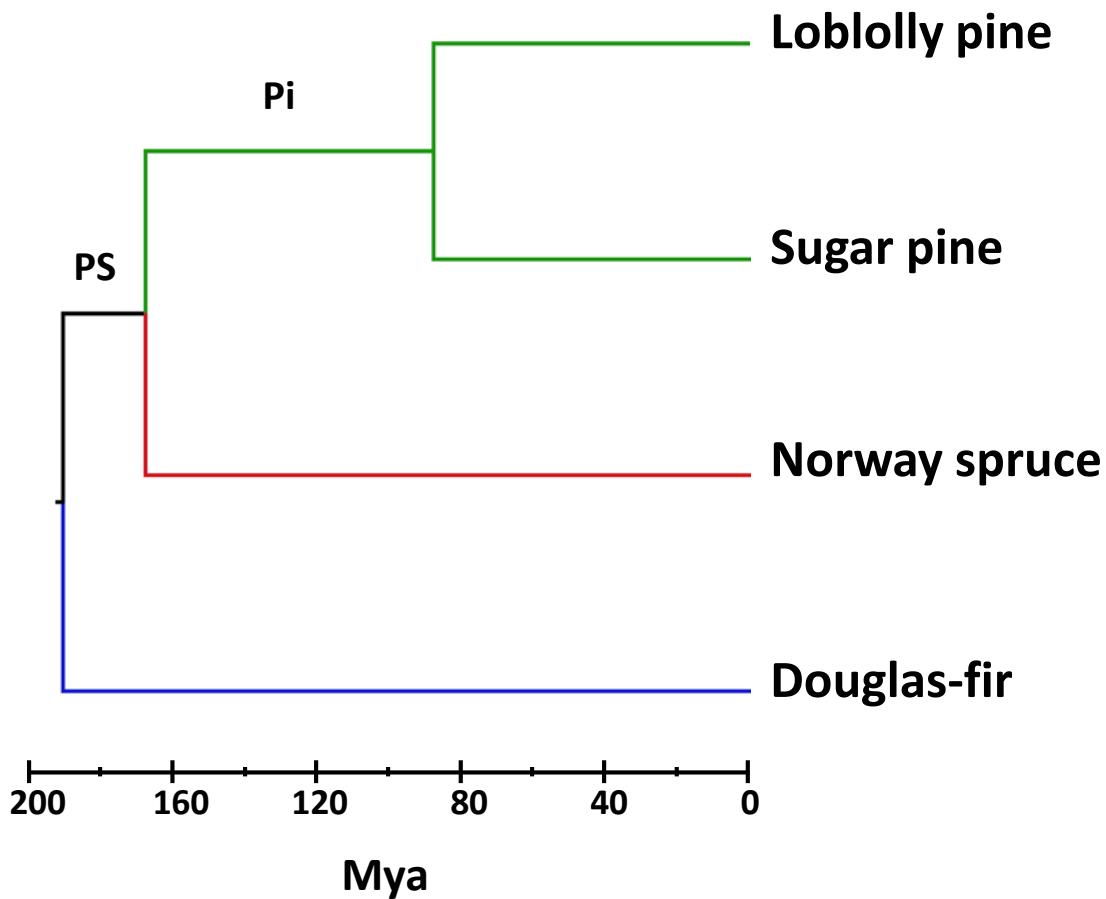
Is there variation in gene turnover between angiosperms and Pinaceae?

# Outline

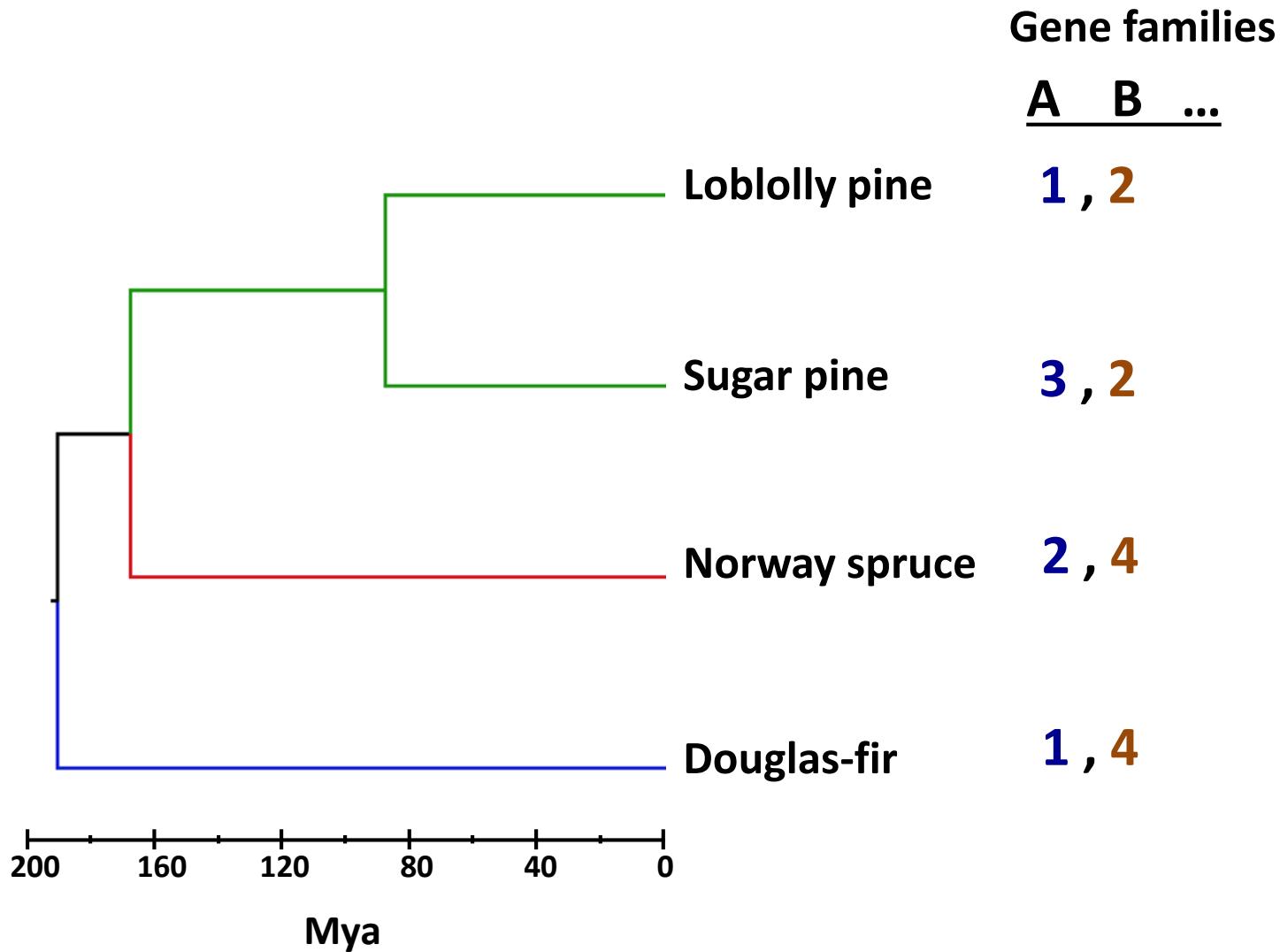
## ◎ Different rates of gene turnover in Pinaceae

1. Issues with gene fragments (false gene duplications)
2. Issues with gene duplications vs. gene losses
3. Issues with species phylogeny

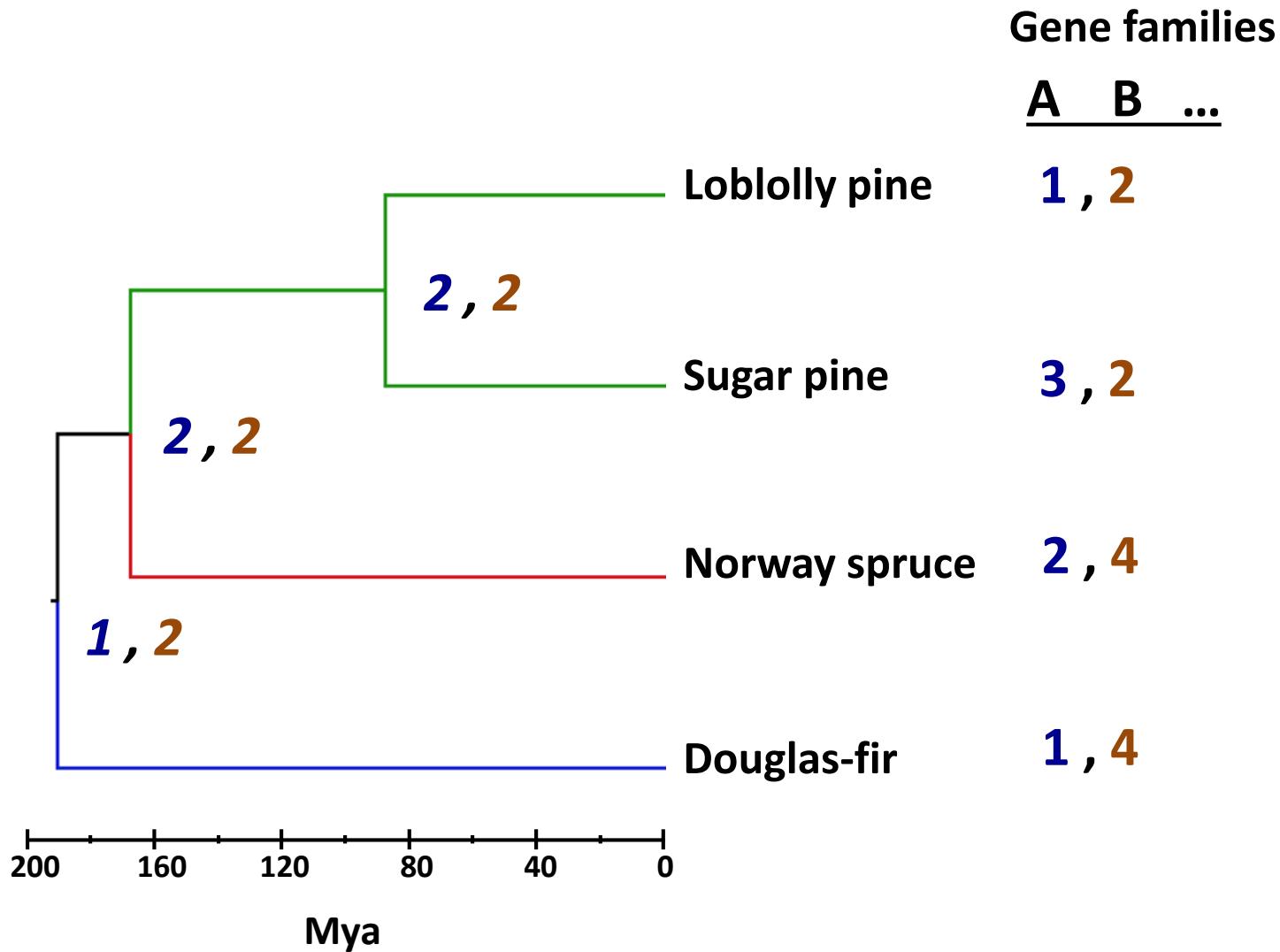
# Gene turnover in four Pinaceae



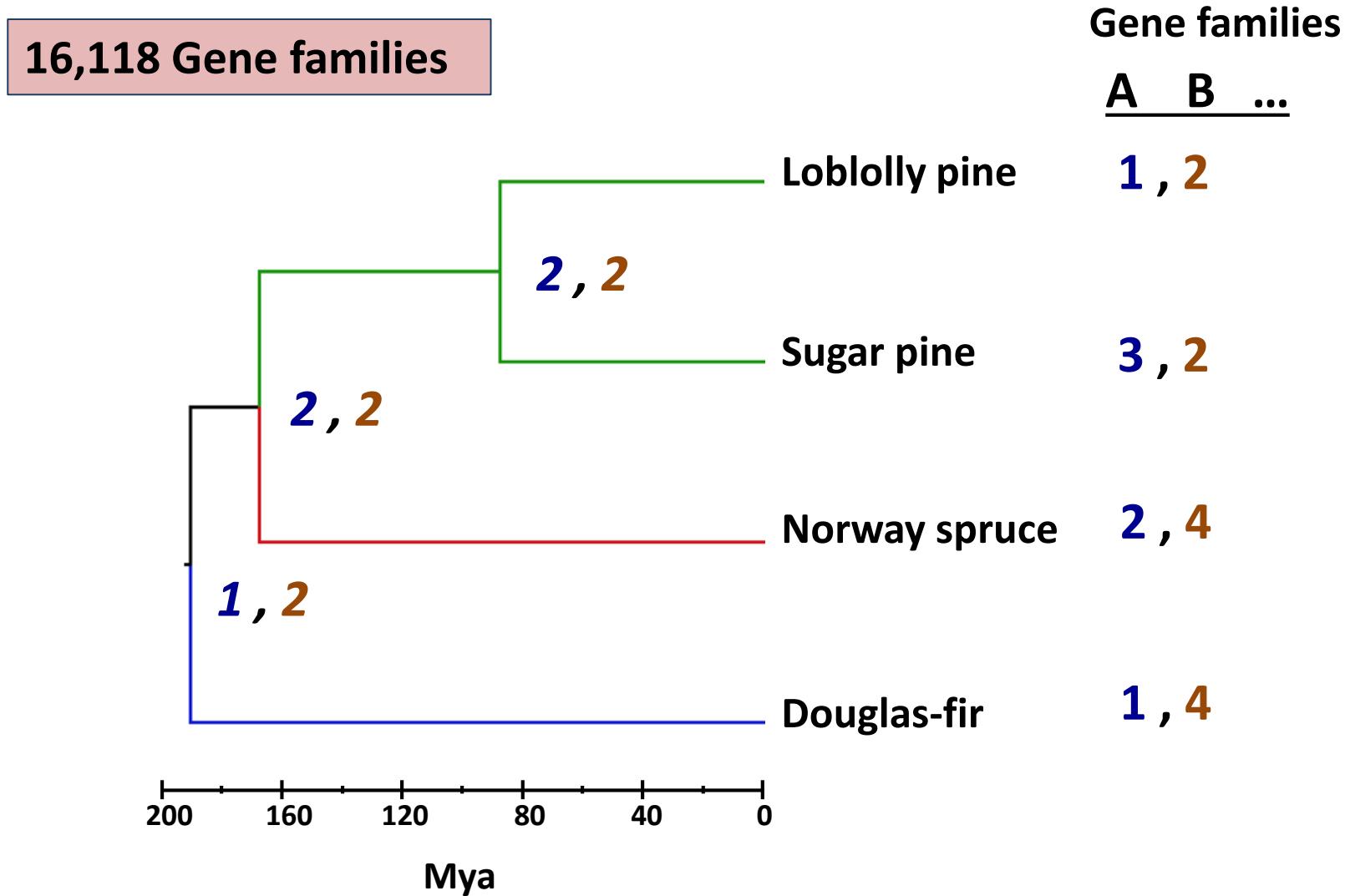
# Observed genes in gene families



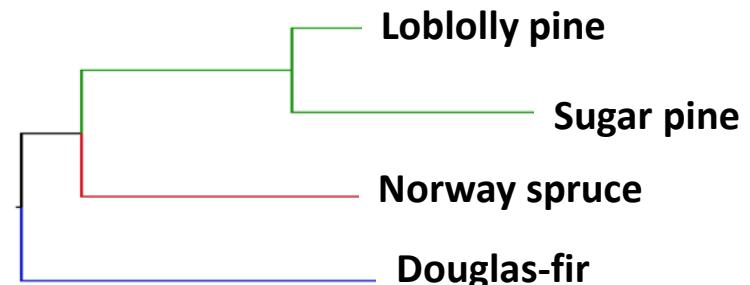
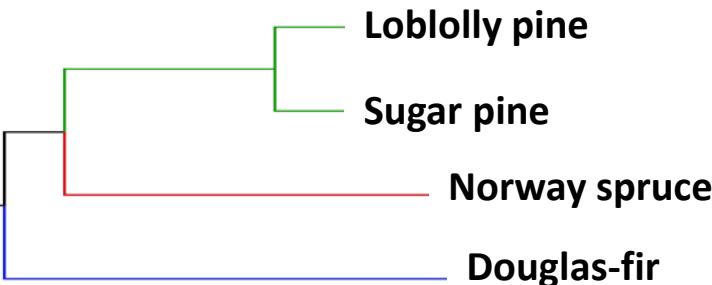
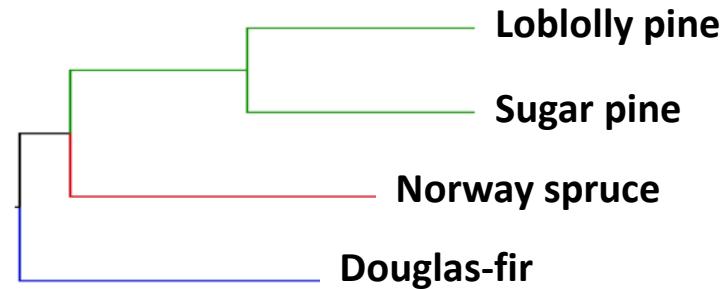
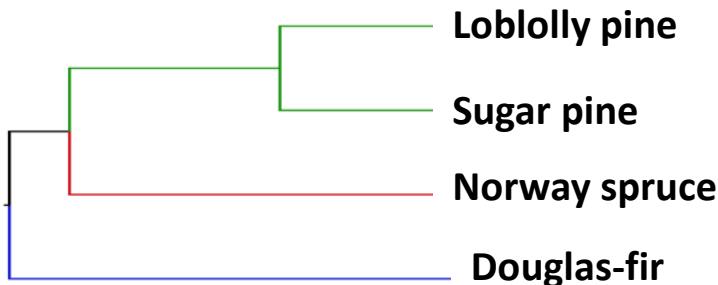
# Estimated ancestral genes in gene families



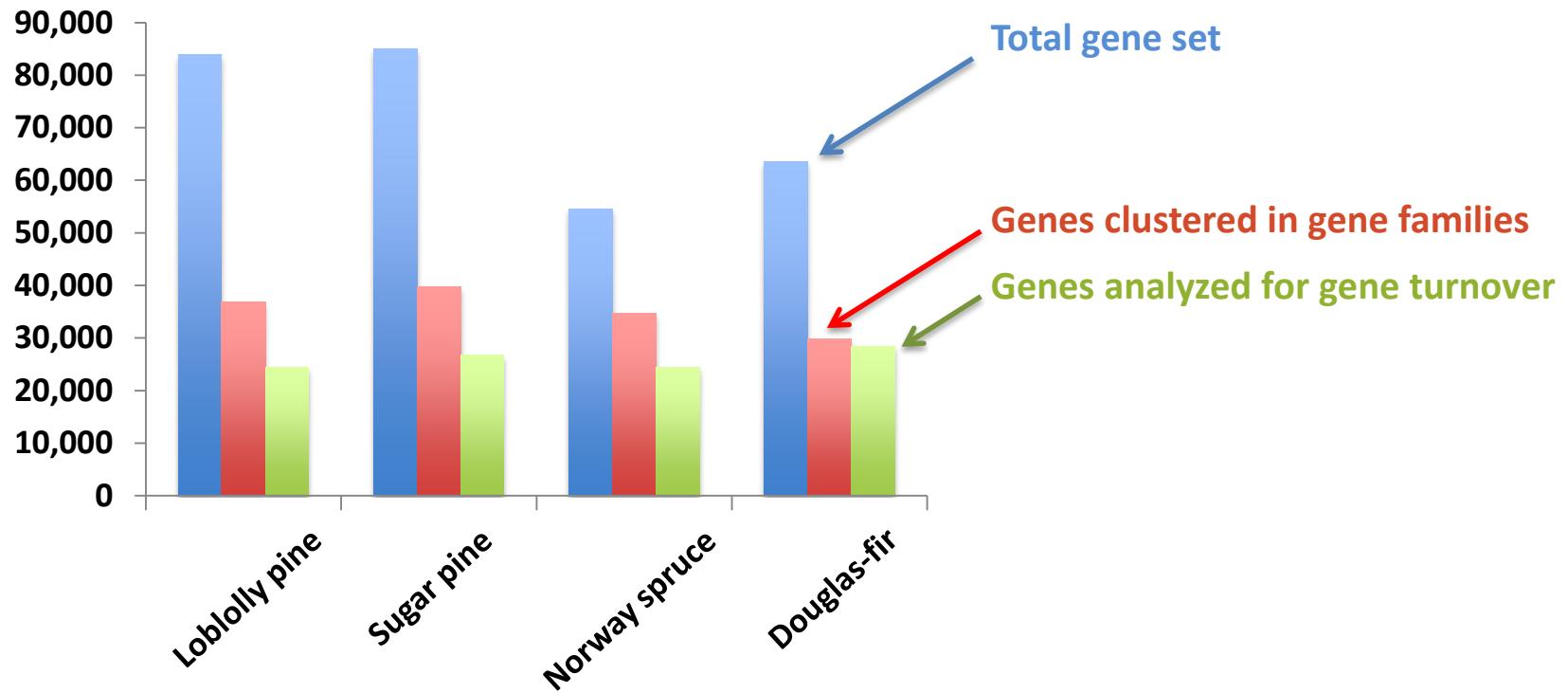
# Estimated ancestral genes in gene families



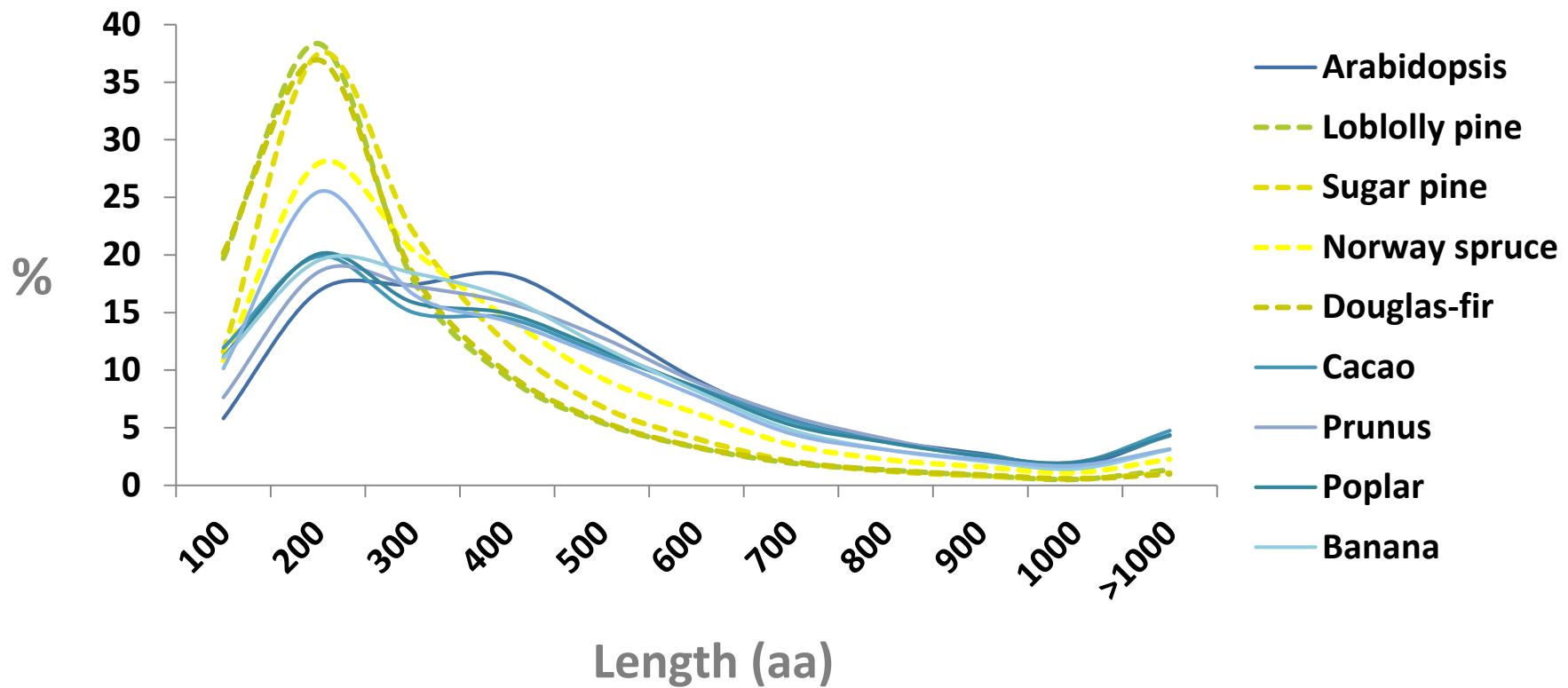
# Different gene turnover rates?



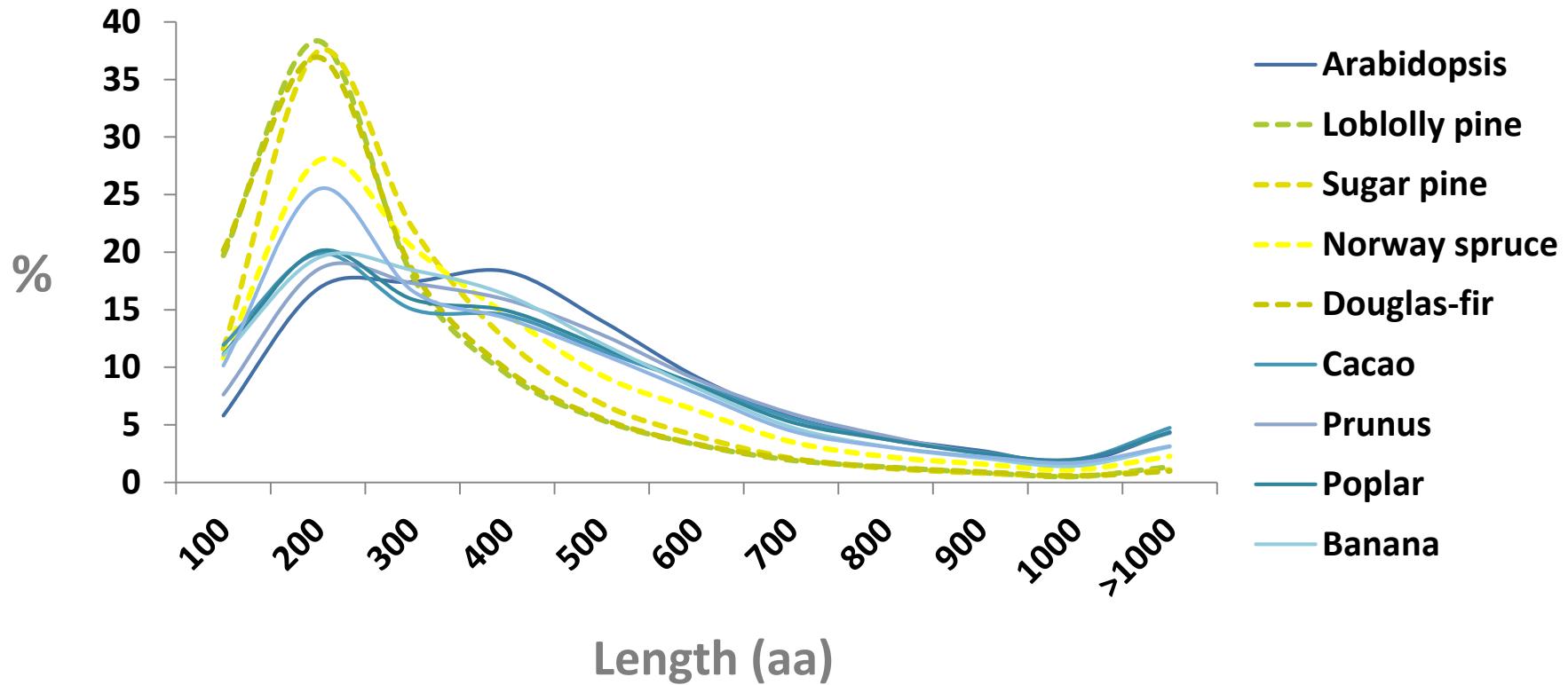
# Gene annotation in Pinaceae



# Gene fragments in Pinaceae



# Gene fragments in Pinaceae

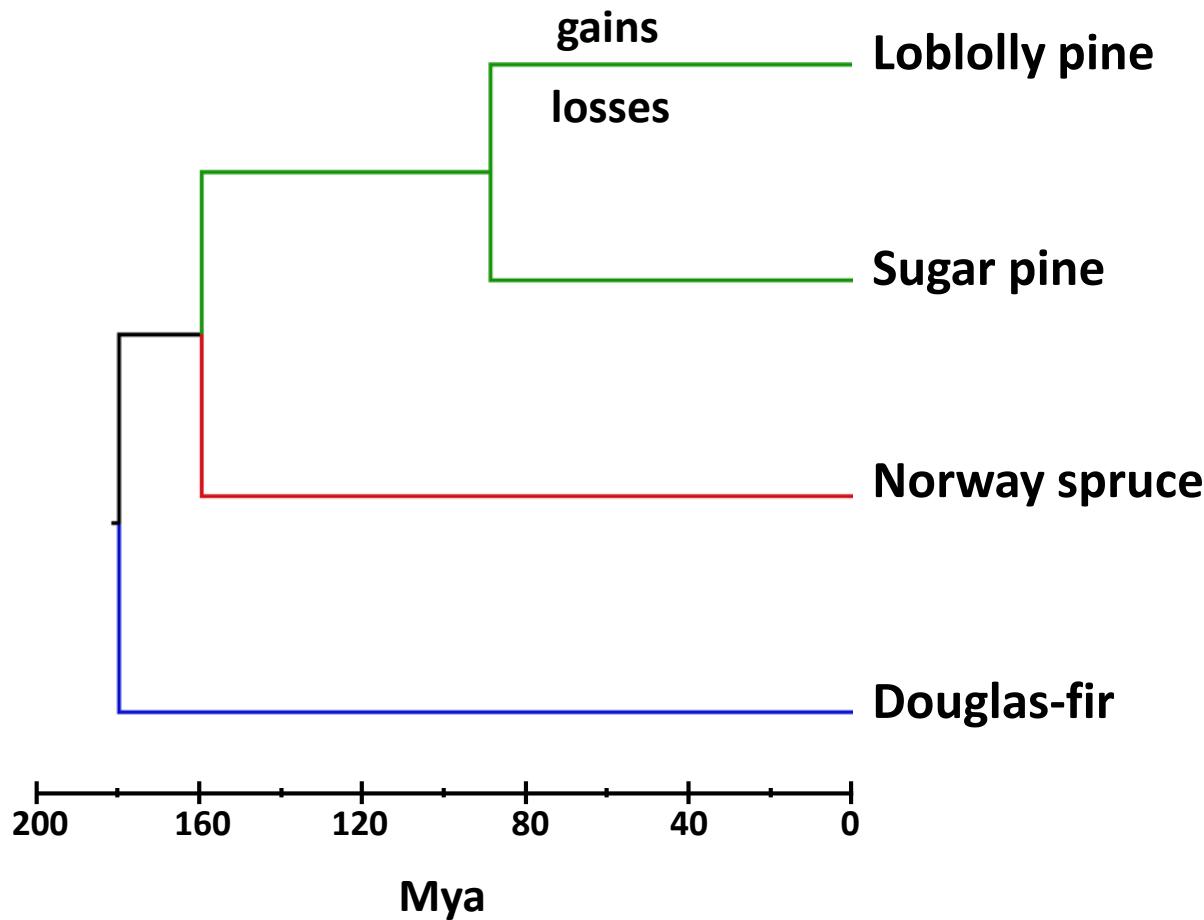


**Complete gene set (CGS):** all genes in gene families included → **104,058 genes**

**Filtered gene set (FGS):** only genes longer than ≥75% of longest gene in the same family included → **75,450 genes**

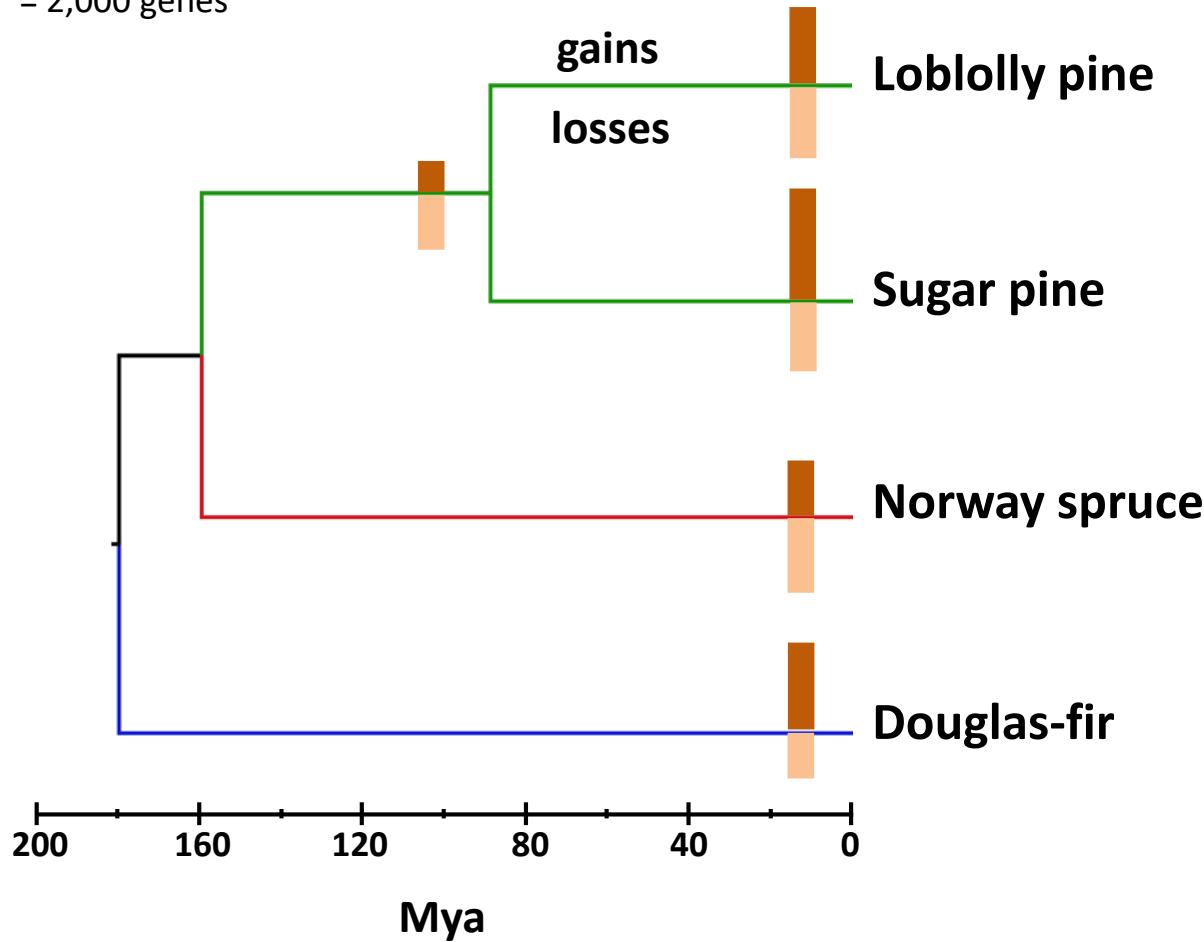
# Gene turnover in four Pinaceae

16,118 Gene families



# Gene turnover in four Pinaceae

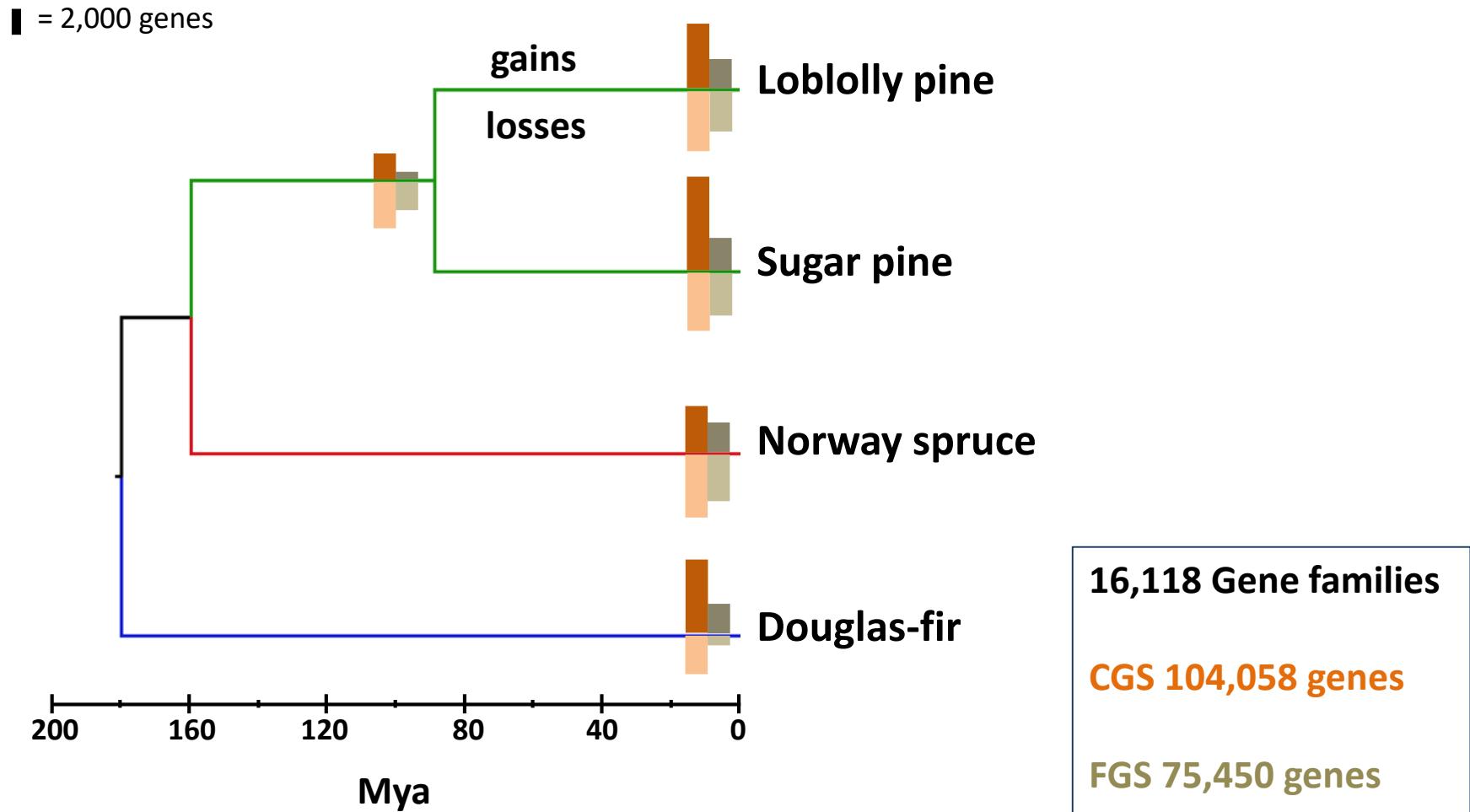
■ = 2,000 genes



16,118 Gene families

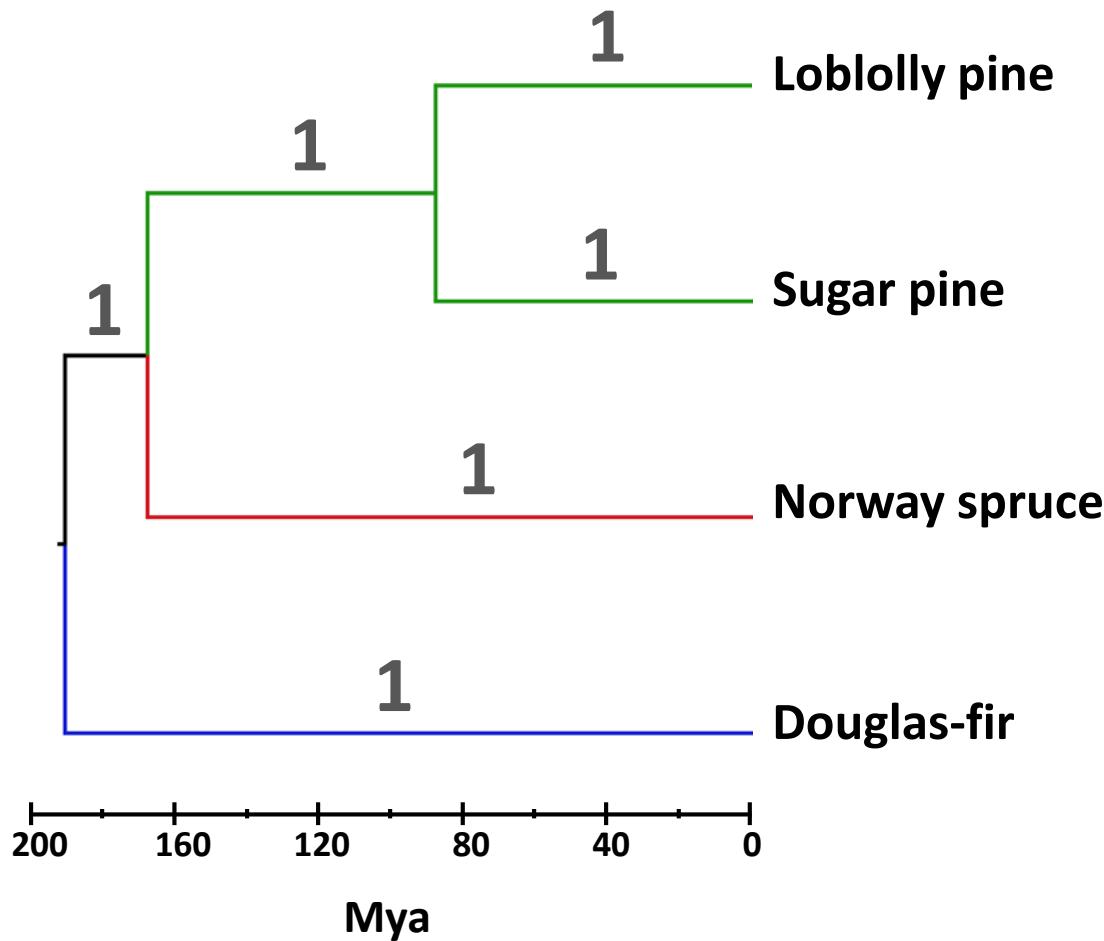
CGS 104,058 genes

# Gene turnover in four Pinaceae

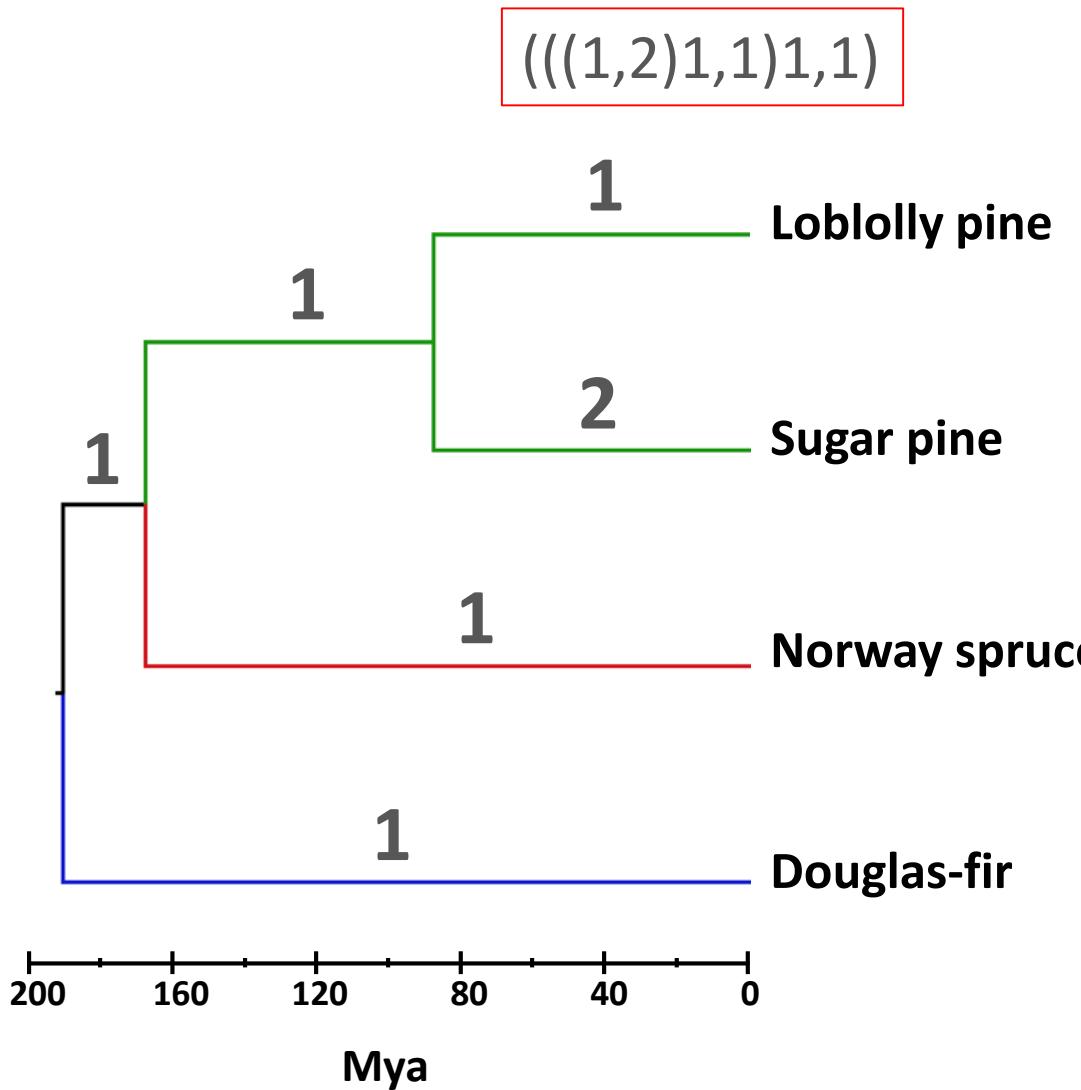


# Models of gene turnover evolution

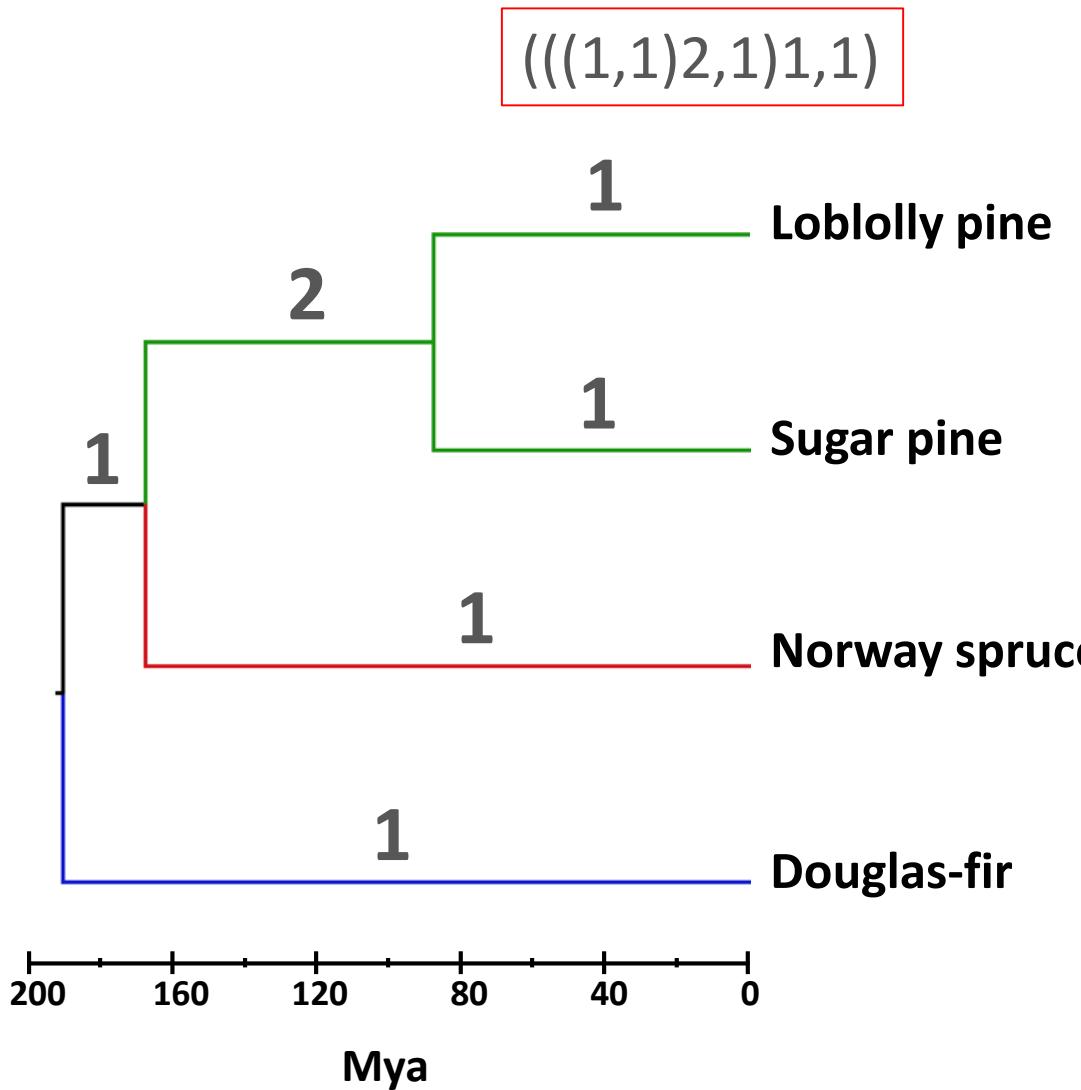
$$(((1,1)1,1)1,1)$$



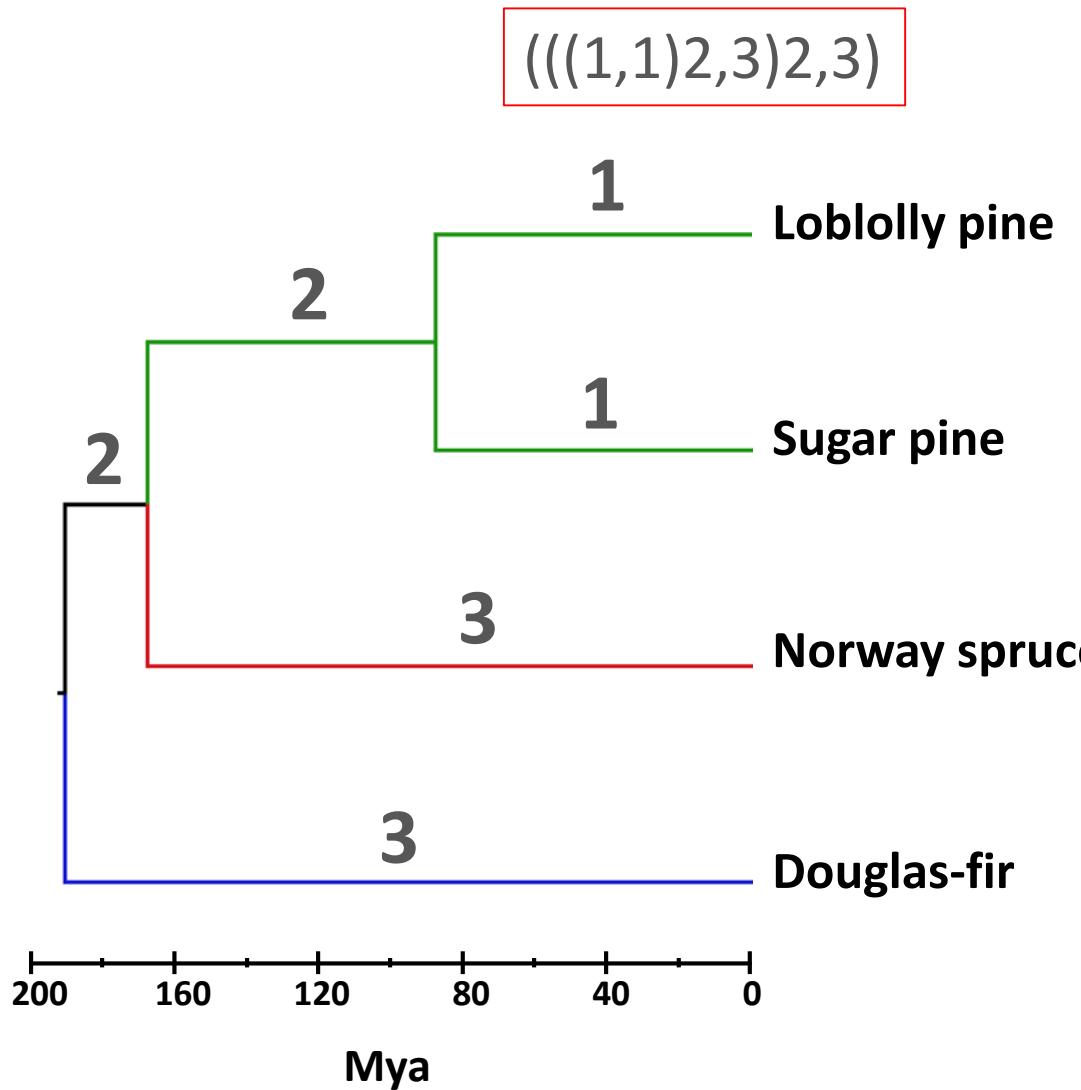
# Models of gene turnover evolution



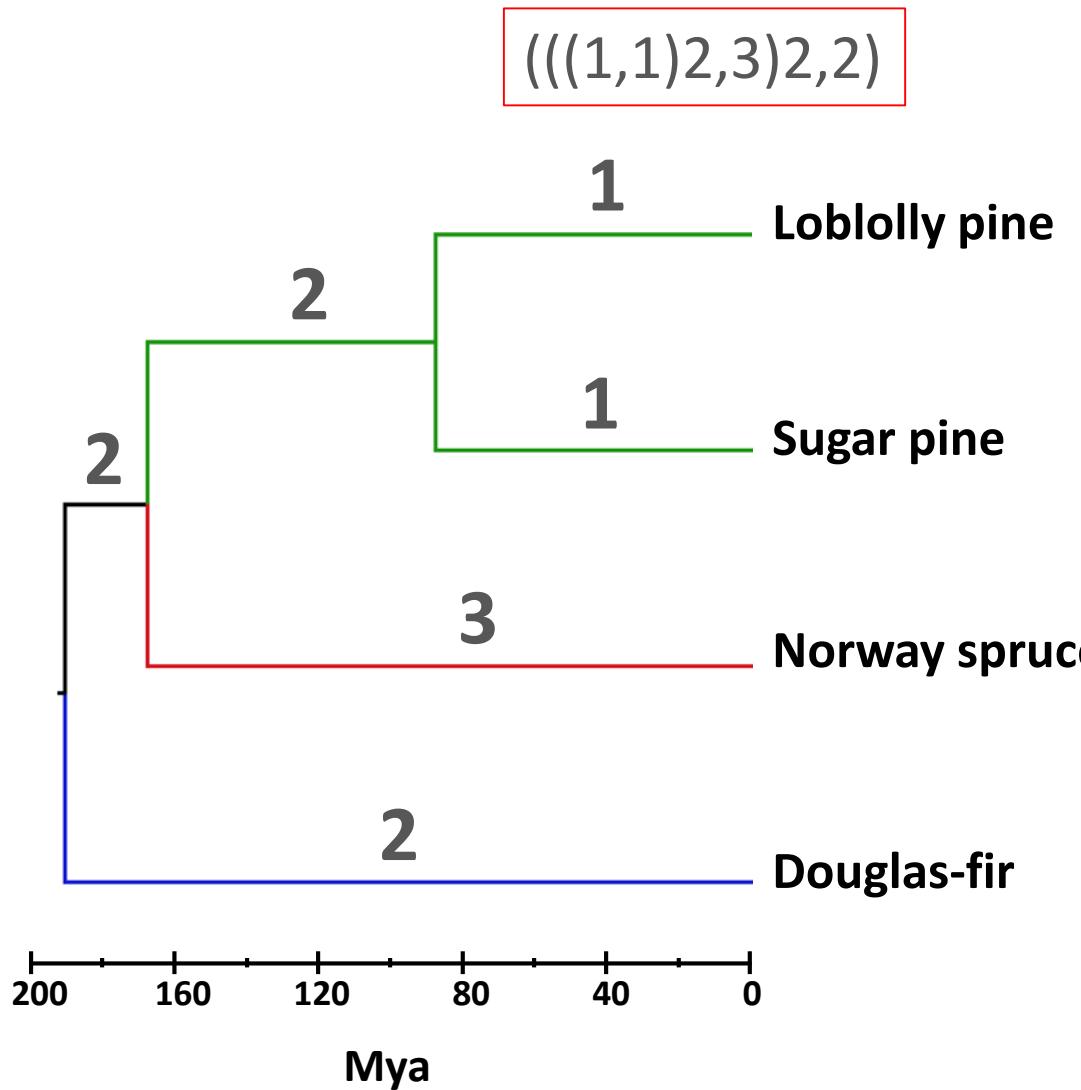
# Models of gene turnover evolution



# Best model for CGS

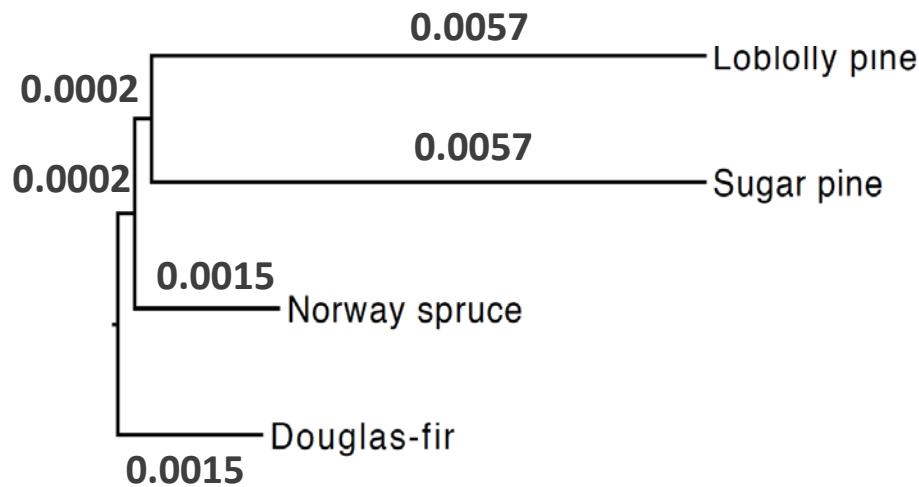


# Best model for FGS

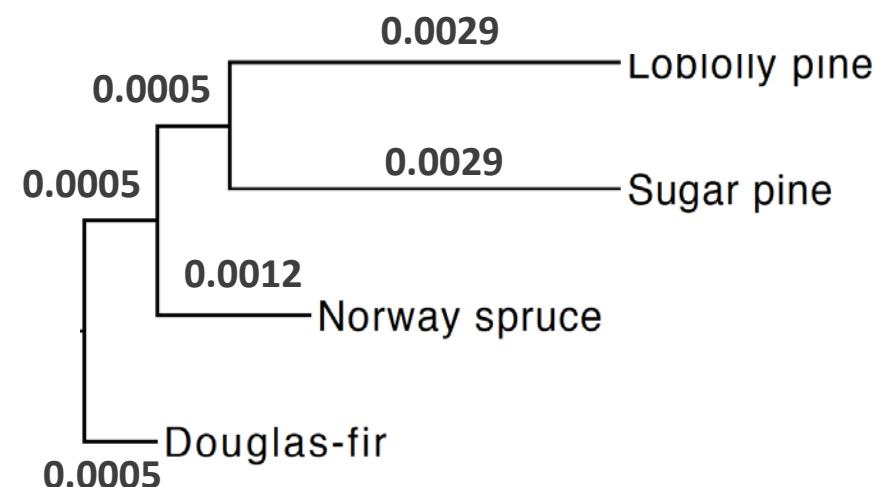


# Higher rates of gene turnover in pine trees for both gene sets

CGS 104,058 genes



FGS 75,450 genes



$((((1,1)2,3)2,3)$

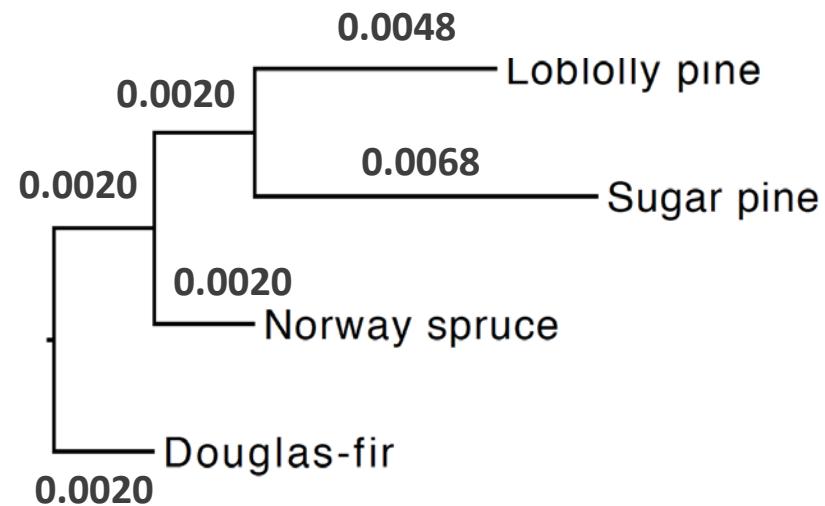
$((((1,1)2,3)2,2)$

~~Issues with gene fragments (false gene duplications)~~

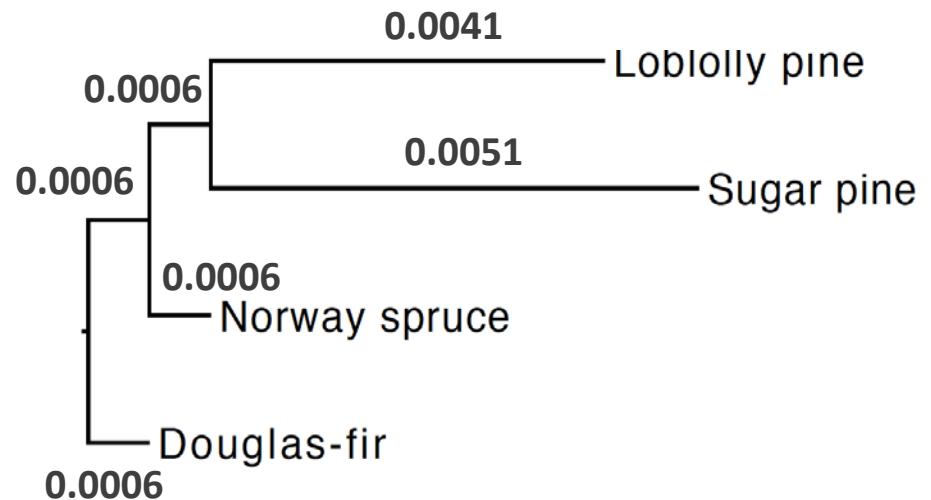
# Removing possible biases due to duplication vs. loss of genes

CGS 104,058 genes

## Gene duplications



## Gene losses



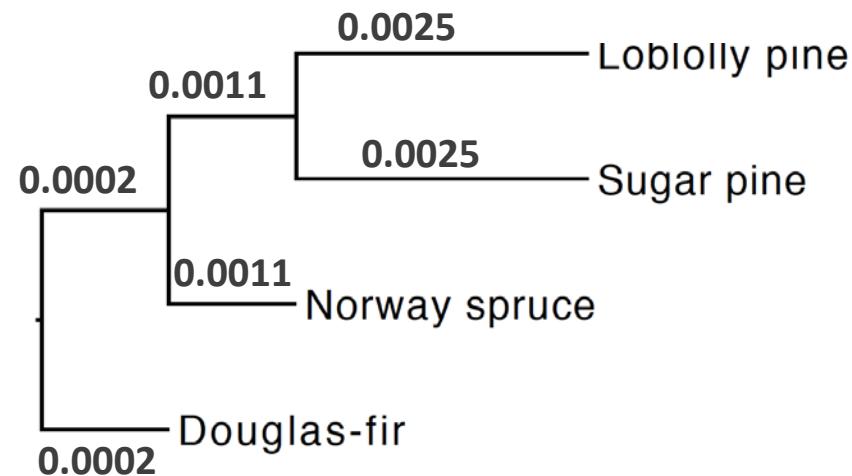
$((1,2)3,3)3,3$

~~Issues with gene duplications vs. gene losses~~

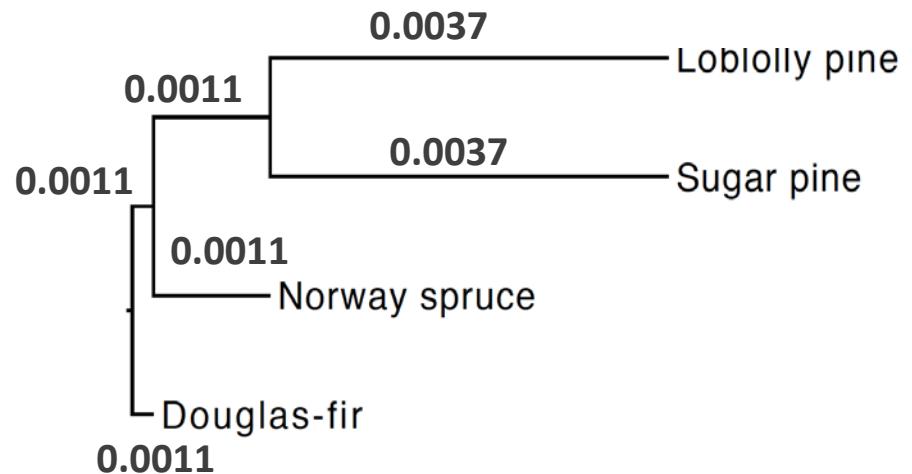
# Removing possible biases due to duplication vs. loss of genes

FGS 75,450 genes

## Gene duplications

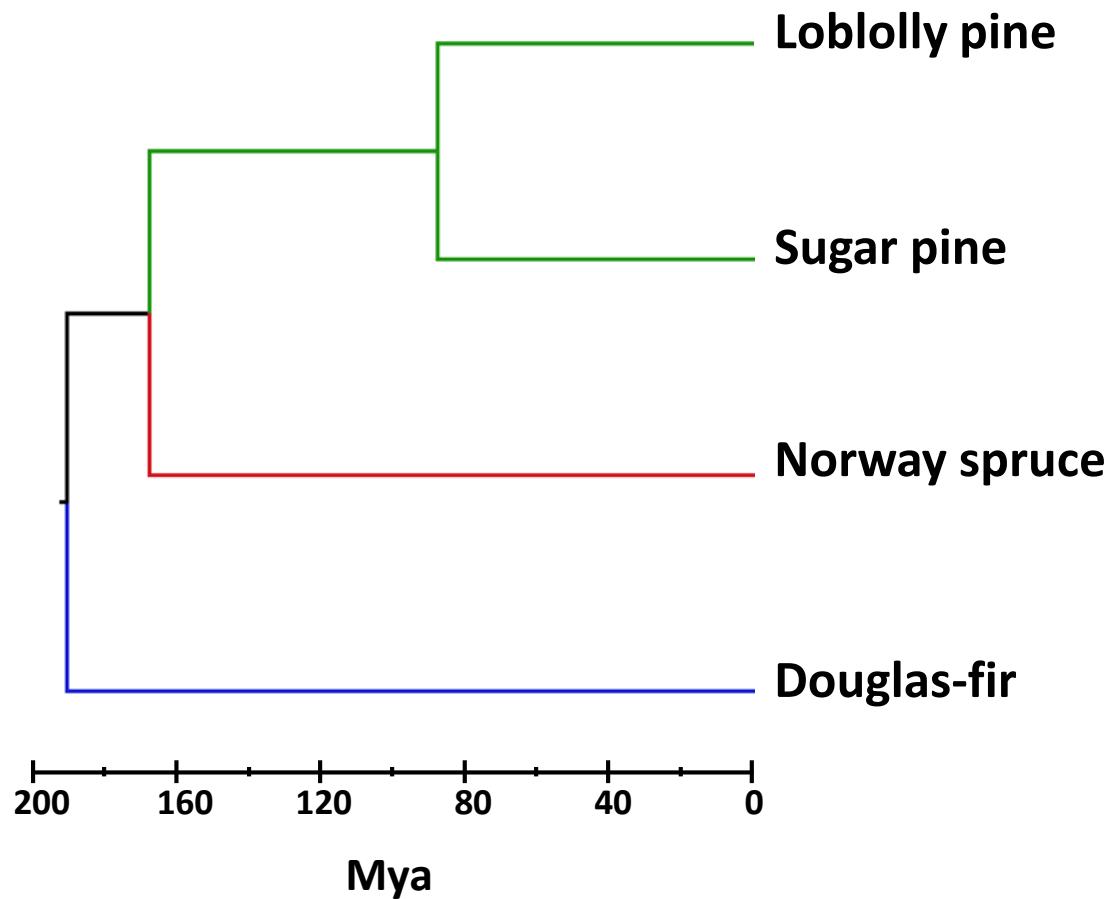


## Gene losses

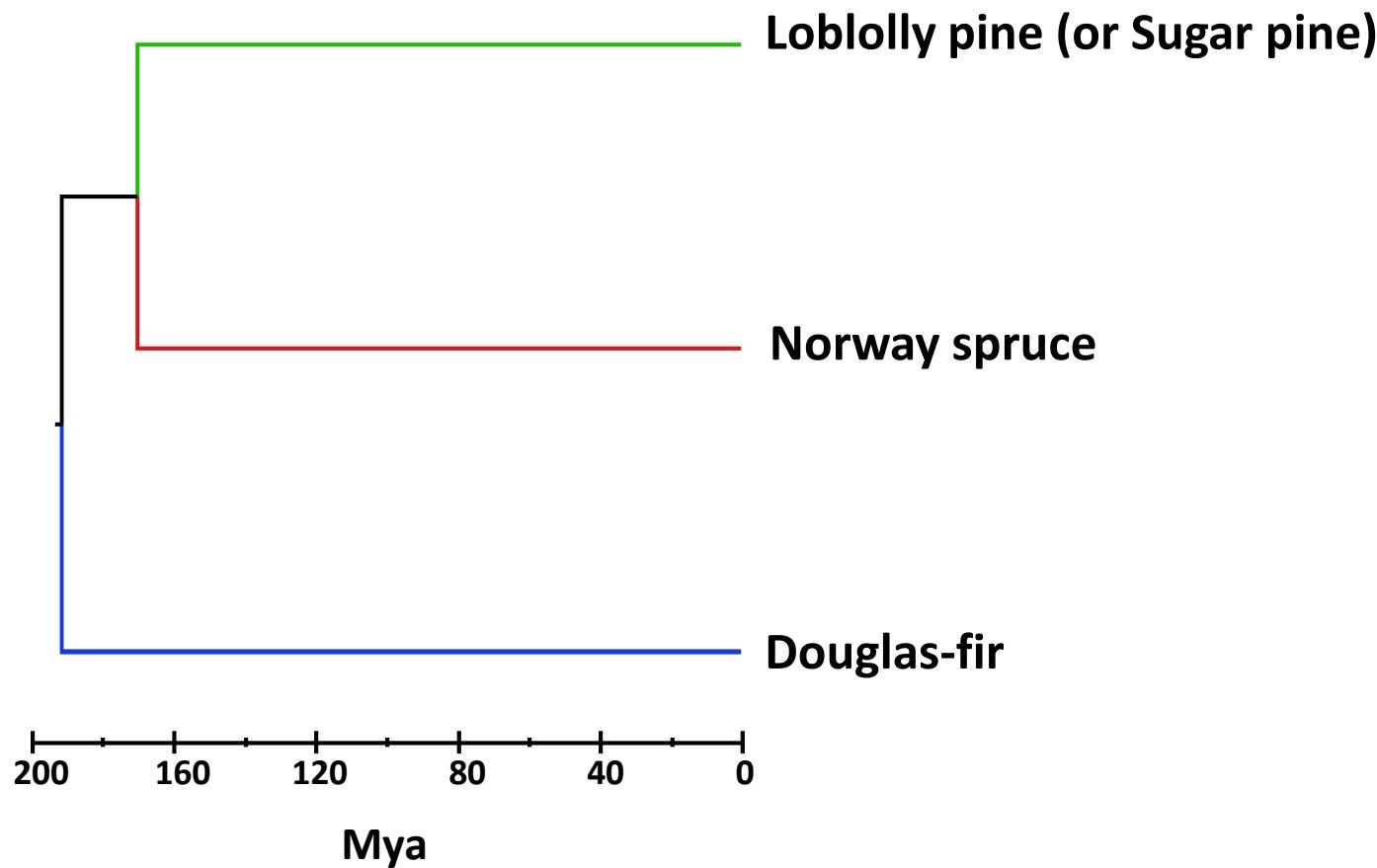


$((((1,1)2,2)3,3)$

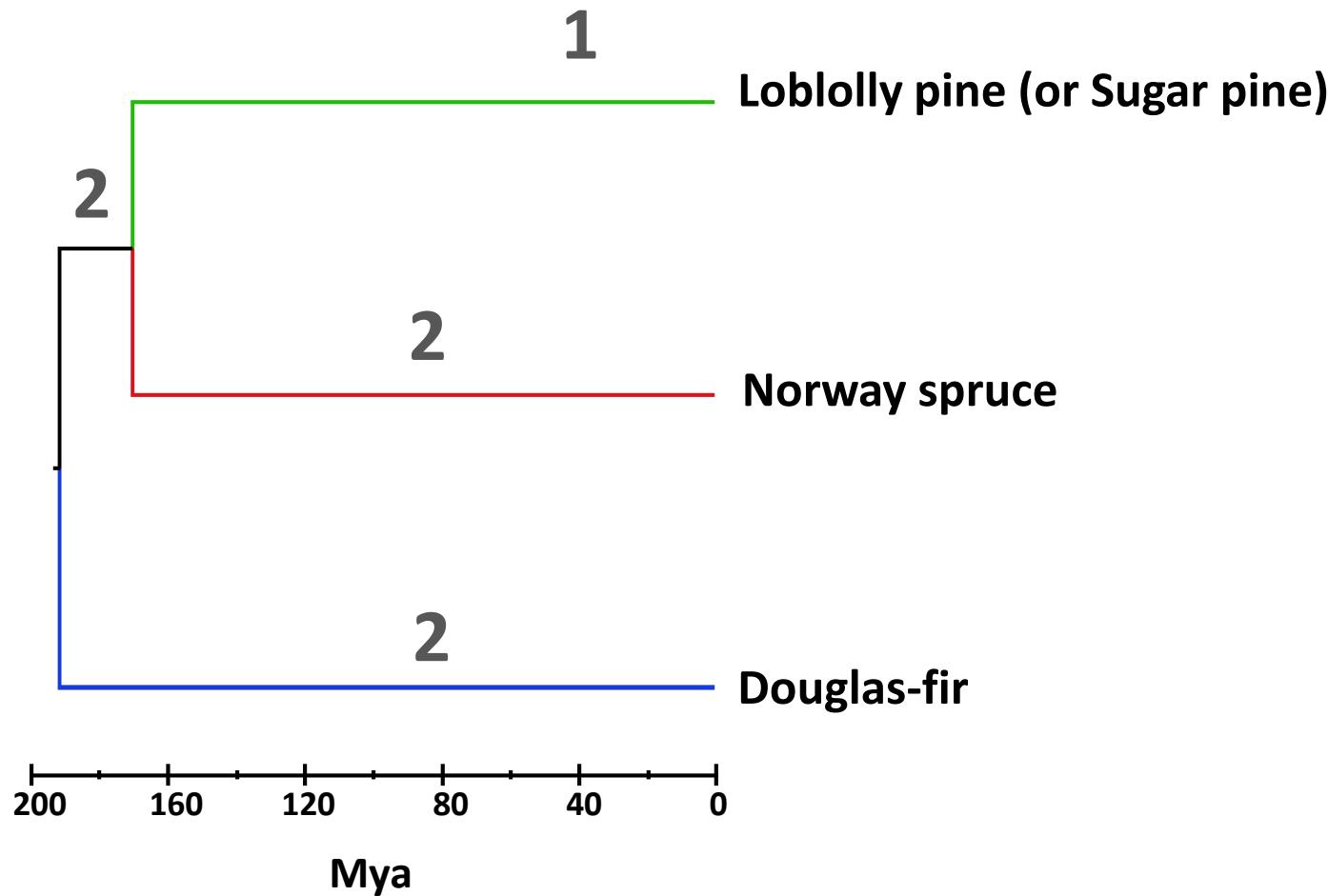
# Removing possible biases due to # of pine trees species



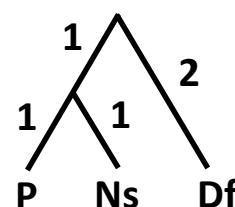
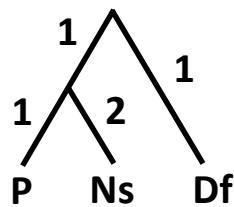
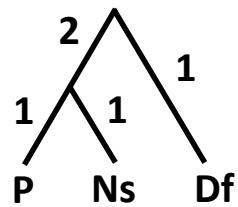
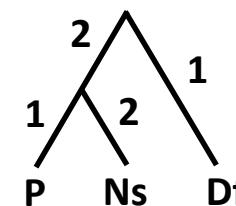
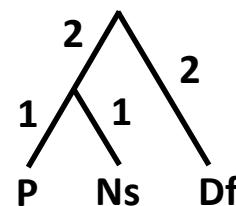
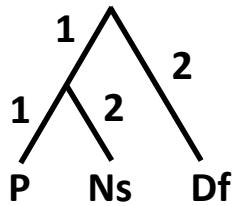
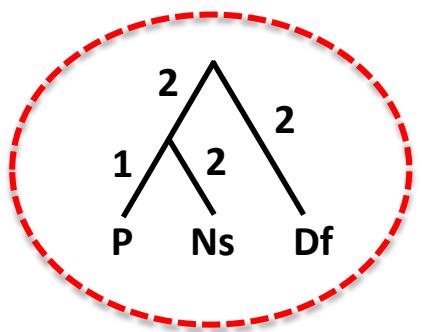
# Removing possible biases due to # of pine trees species



# Removing possible biases due to # of pine trees species

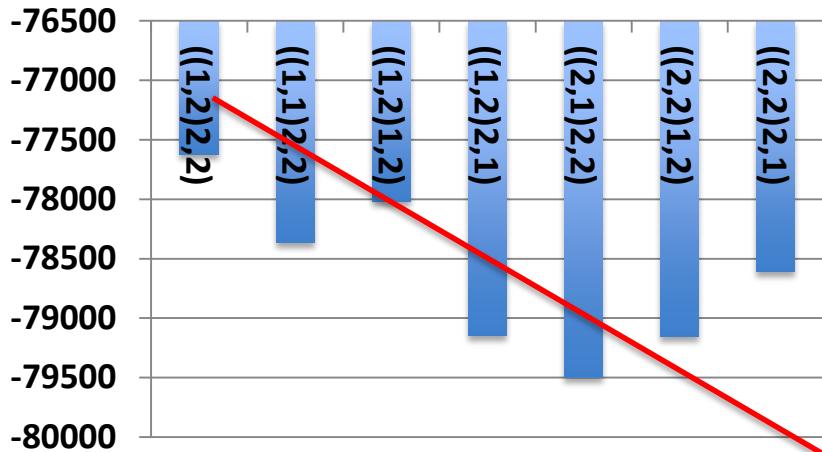


# Removing possible biases due to # of pine trees species



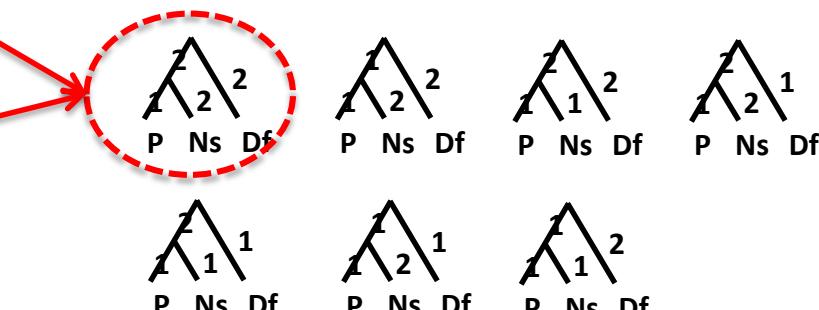
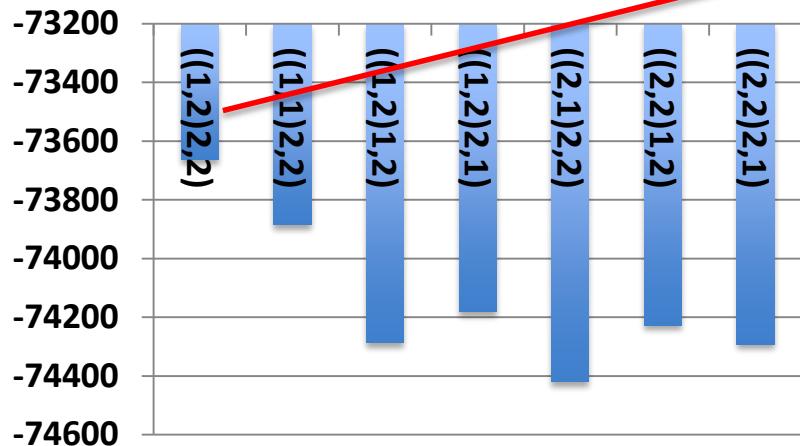
P: pine tree. Ns: Norway spruce. Df: Douglas-fir

# A ((Loblolly pine,Spruce)PS,DF)



	$2-\lambda\ ((1,2)2,2)$	$2-\lambda\ ((1,2)2,2)$
Loblolly pine	0.0027	na
Sugar pine	na	0.0036
Pines (Pi)	na	na
Norway spruce (Ns)	0.0013	0.0013
Pines+spruce (PS)	0.0013	0.0013
Douglas-fir (Df)	0.0013	0.0013

# B ((Sugar pine,Spruce)PS,DF)



Issues with species phylogeny

# Outline

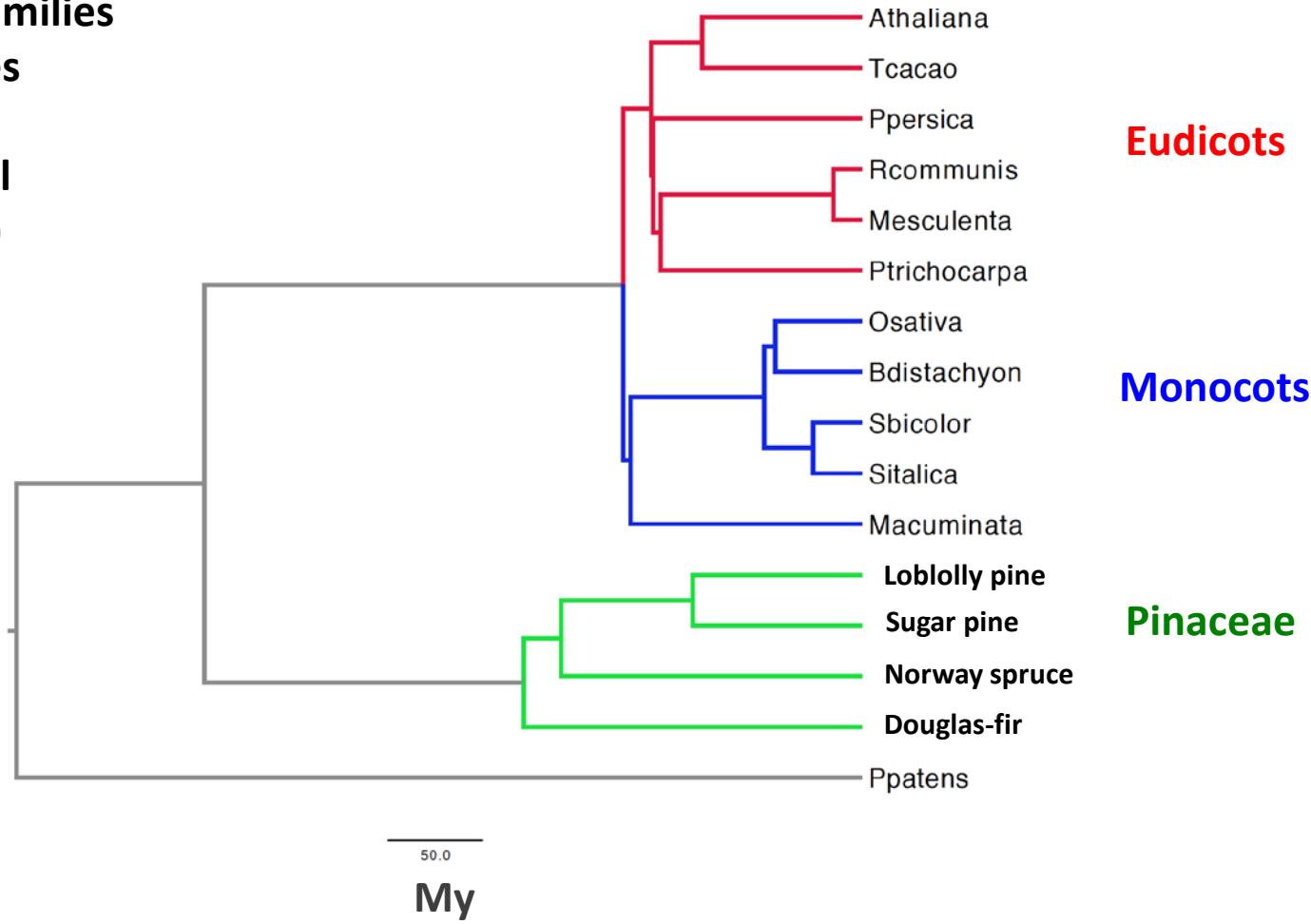
## ● Rates of gene turnover in Pinaceae vs. angiosperms

Is there variation in gene turnover between angiosperms and Pinaceae?

# Gene turnover in seed plants

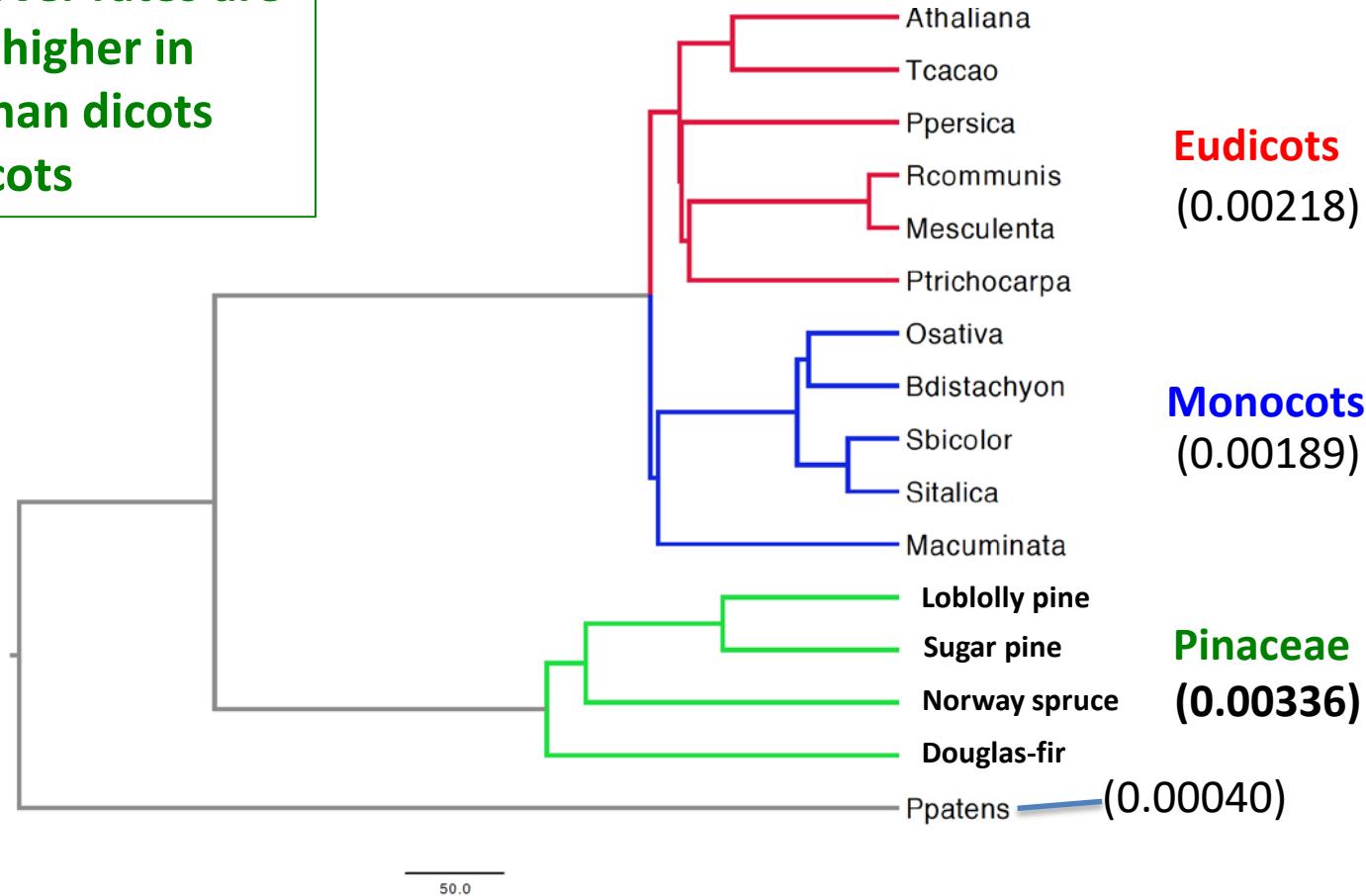
6,459 gene families  
230,055 genes

4-rates model  
(color-coded)



# Gene turnover in Pinaceae is higher than other eukaryotes

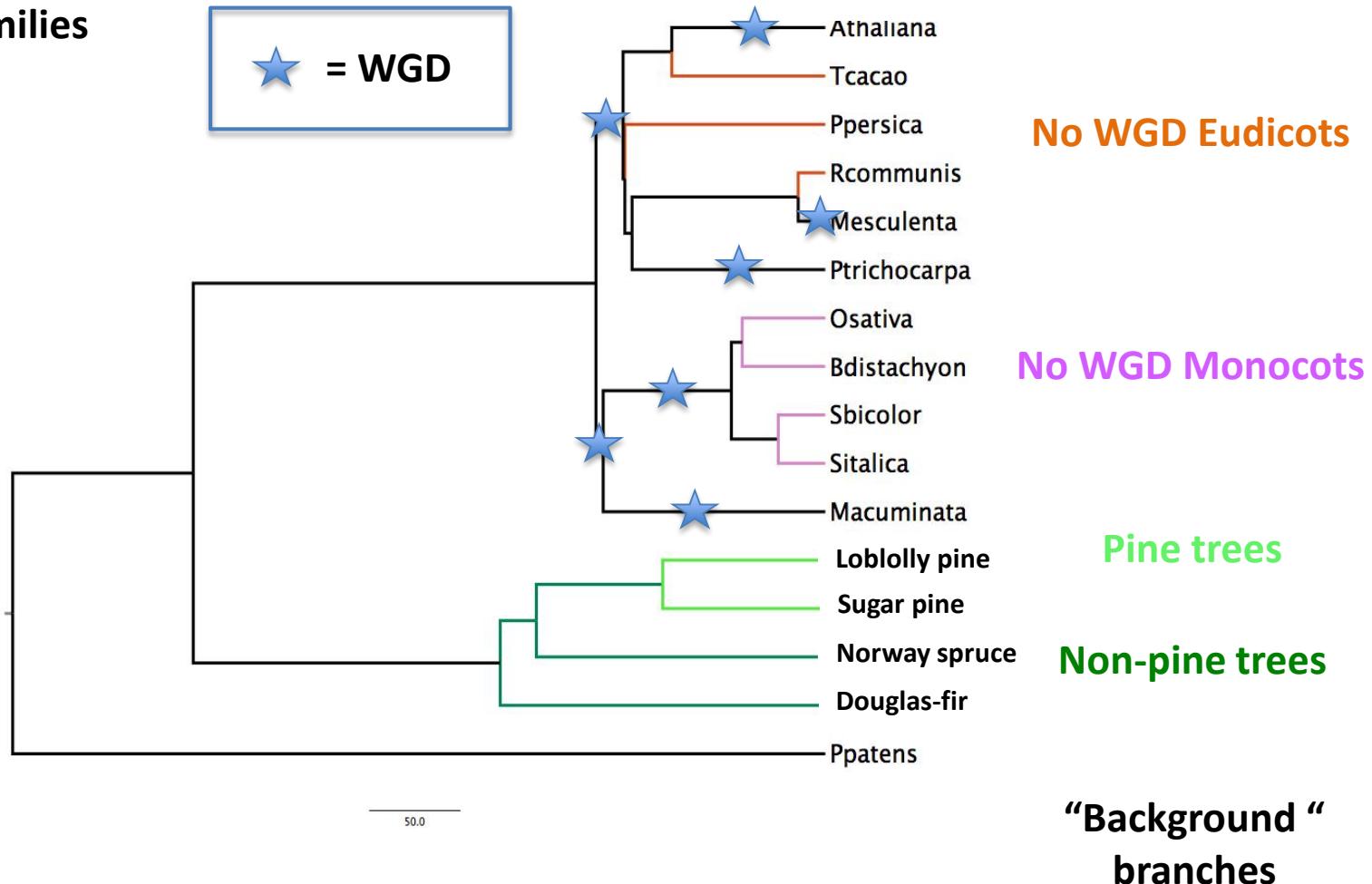
Gene turnover rates are  
1.54-1.78x higher in  
Pinaceae than dicots  
and monocots



# Gene turnover in Pine trees is higher than in angiosperm lineages with no Whole-Genome Duplication (WGD)

6,459 gene families  
230,055 genes

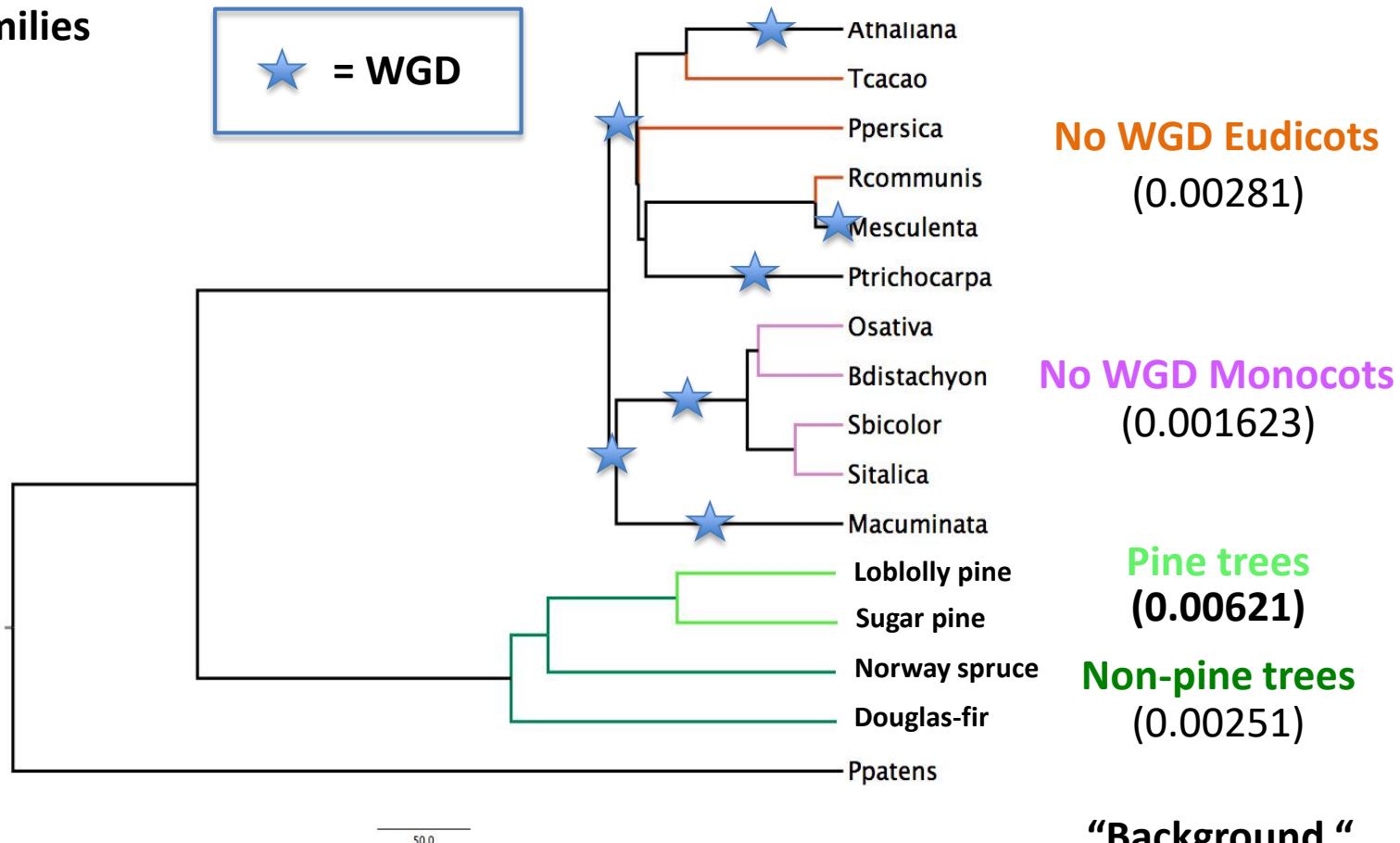
5-rates model  
(color-coded)



# Gene turnover in Pine trees is higher than in angiosperm lineages with no Whole-Genome Duplication (WGD)

6,459 gene families  
230,055 genes

5-rates model  
(color-coded)



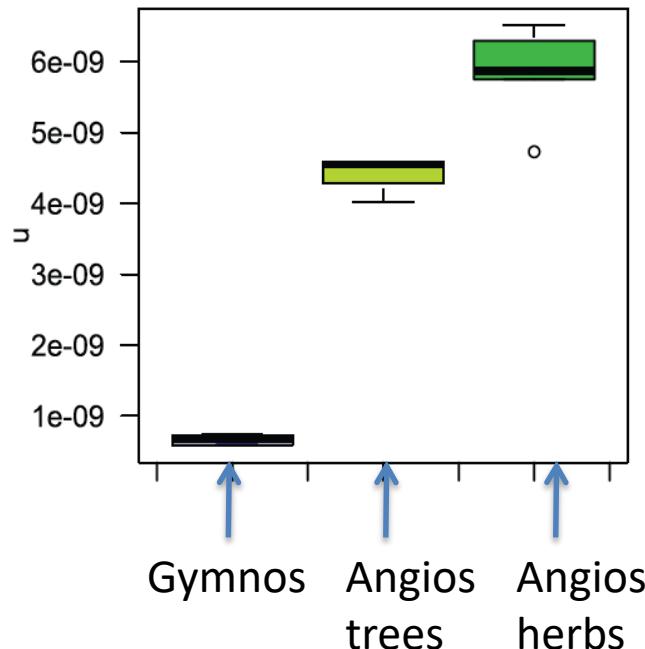
“Background “  
branches  
(0.00129)

# Contrasting Rates of Molecular Evolution and Patterns of Selection among Gymnosperms and Flowering Plants

Amanda R. De La Torre,<sup>\*,1,2</sup> Zhen Li,<sup>3,4</sup> Yves Van de Peer,<sup>3,4,5</sup> and Pär K. Ingvarsson<sup>2,6</sup>

**Gymnosperm rates of molecular evolution per unit time are, on average, seven times lower than angiosperm species**

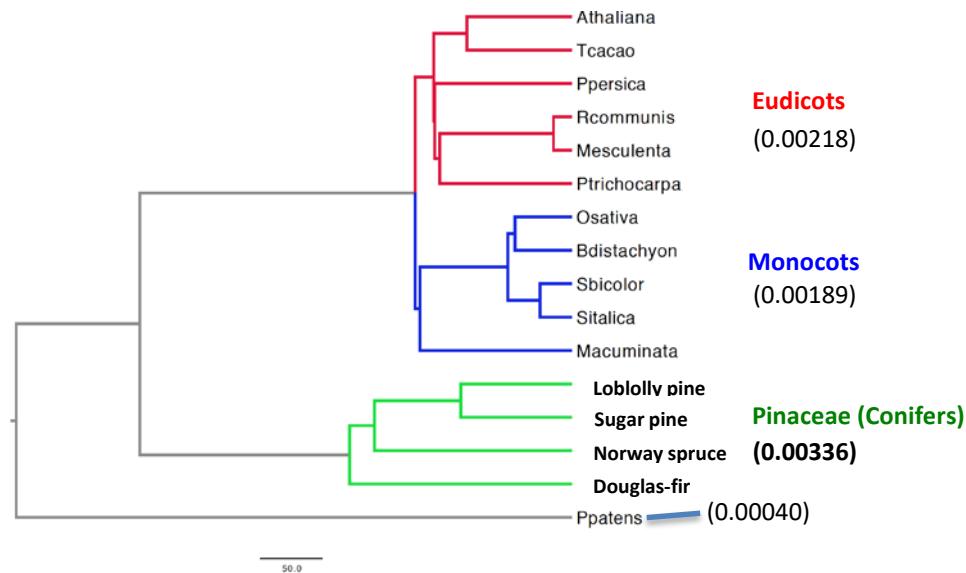
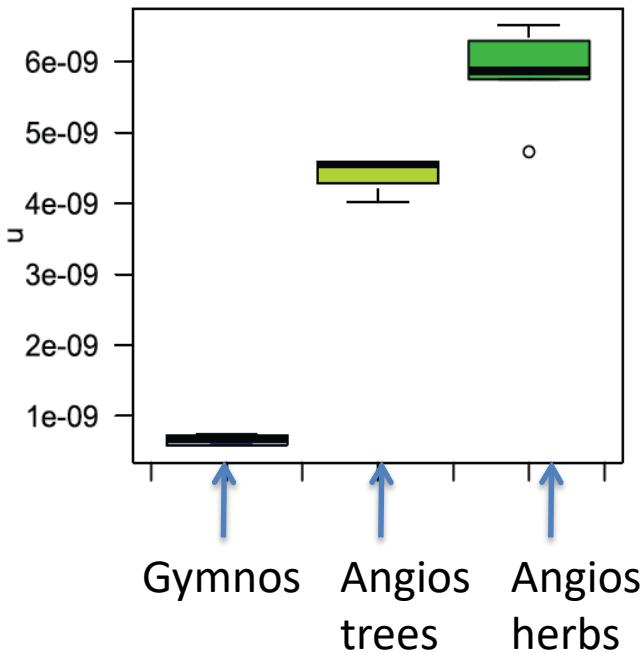
“Longer generation times and larger genome sizes are some of the factors explaining the slow rates of molecular evolution found in gymnosperms.”



# Contrasting Rates of Molecular Evolution and Patterns of Selection among Gymnosperms and Flowering Plants

Amanda R. De La Torre,<sup>\*,1,2</sup> Zhen Li,<sup>3,4</sup> Yves Van de Peer,<sup>3,4,5</sup> and Pär K. Ingvarsson<sup>2,6</sup>

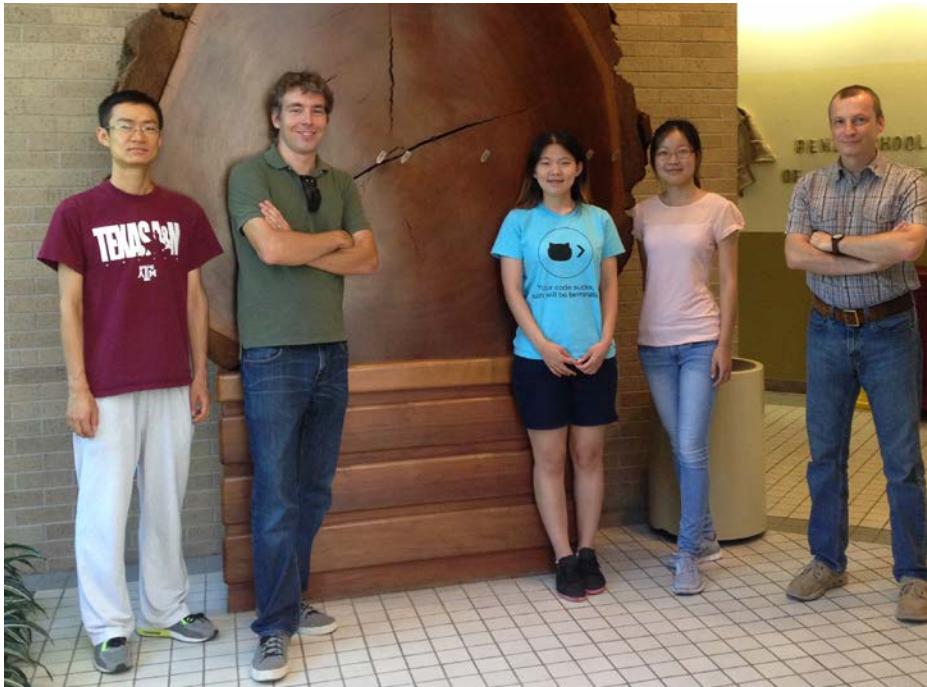
## *Contrasting rates of gene turnover and nucleotide substitutions among Pinaceae and flowering plants*



# Conclusions

- ❖ Accelerated gene turnover in pine trees vs. two other Pinaceae
- ❖ This pattern is robust to potential biases due to presence of gene fragments, gene duplications vs. gene losses and presence of multiple pine trees in the phylogeny
- ❖ Gene turnover may be accelerated in conifers as opposed to angiosperm lineages that did not experience whole-genome duplications

# Acknowledgments



## *Collaborators*

**Jill Wegrzyn.** University of Connecticut  
**Carol Loopstra.** Texas A&M University  
**Michelle Lawing.** Texas A&M University

## *Lab members*

**Tomasz Koralewski**  
**Weixi Zhu**  
**William McCoy**  
**Xuan Lin**  
**Jingjia Li**

